

Our emperors have no clothes

The ITER fusion project demonstrates a solidity of purpose that is sorely lacking across the rest of the energy research spectrum.

Last week's formal agreement in Paris to begin construction of the ITER experimental fusion reactor at Cadarache in France was a satisfying culmination of decades of effort by the fusion community. The project may or may not ensure the technical future of magnetic fusion energy, but it will certainly provide an invaluable focal point for its investigation for decades to come.

The agreement to build ITER — which, in a landmark move, includes significant contributions from both India and China — is out of character, however. There is no corresponding effort, collaborative or otherwise, to do the research that is needed to aggressively pursue any of the other multifarious approaches to tackling the mounting global energy crisis.

Energy research and development has actually collapsed since the aftermath of the last energy crisis in the 1970s. Indeed, according to an assessment by the Pacific Northwest National Laboratory in Washington state, global expenditure on it has fallen by two-thirds since 1974. This time, climate change and a vast upsurge in the Asian demand for energy, as well as a crisis in the Middle East, mean it is almost universally acknowledged that the energy crisis is here to stay. But there is still no serious effort to revive energy research.

The reasons (or excuses) for inaction vary from place to place. One universal theme lurking in the background is the belief that market forces will pull viable energy sources to fruition, without the need for publicly supported research. Oddly enough, this argument doesn't stop governments supporting basic and applied research in areas such as public health, where they perceive a pressing need to push technology forward into the market. In any case, market forces are largely absent, as most users of fossil fuel still don't pay the real cost of their carbon emissions.

Energy research

It is true that the short-term economic feasibility of renewable energy sources, for example, or of clean-burning coal-fired power stations, will be determined largely by market conditions. However, in the longer term, radical breakthroughs in everything from photovoltaic materials to the efficiency of cars and trucks are critically important to the prospects for different approaches to energy production, transmission and use.

Over the past 30 years, only Japan, which is dependent on imported fossil fuels, has maintained energy research at a reasonably high level. And even Japan's programme is seriously skewed in favour of approaches to nuclear fission that look increasingly unlikely to meet its energy needs.

In the United States, energy research at least has a natural home, in the Department of Energy. The department's research programmes have some notable strengths: its basic energy-sciences programme, for example, supports a raft of basic research projects and facilities in fields, such as materials science, that may underpin future progress in

energy technologies. Its network of national laboratories provides a good interdisciplinary foundation for both basic and applied energy research. And its research efforts in energy conservation have grown over the past decade. However, its investment in research on renewable energy remains tiny, especially compared with the levels that were briefly attained under Jimmy Carter's presidency. On energy policy, as on other matters, Carter's legacy continues to improve with age. If the new Democratic majority in Congress is serious about energy policy, it will do its best to restore these programmes.

Underfunding and neglect

In Europe, energy research programmes have been eviscerated since the 1970s, despite political leaders' claims to be taking climate change seriously. The nuclear fission programmes that dominated at that time have been abandoned (except in France), and nothing has been established in their place. Applied energy research has no natural institutional home and no clear political constituency (in the shape of a group of dependent institutions or individuals who will fight for its budget). With the exception of fusion and ITER, the overall European effort is paltry, and is mainly based in universities — which are good at identifying ideas, but are not equipped to take them forward into commercial application.

Some proposals have been developed to confront this. Germany is spending €50 million (US\$65 million) each year at the European Institute for Energy Research at Karlsruhe, and Britain is planning to spend a similar amount (matched by private funds) at an Energy Technology Institute, to be established in 2008. The seventh European Framework Programme, which starts next year, identifies energy as one of nine priorities (transportation is another), but devotes less than 5% of its total resources — €2.3 billion over six years — to its pursuit.

None of this is commensurate with the immense challenges of energy security and climate change facing the world. The ITER project, admirable though it is, should be merely a component of a concerted effort. Fusion is decades away from fruition, and even its proponents don't expect it to contribute to the world's need for alternatives to fossil fuel within the next 30 years.

Fusion is also unusual in that its prototyping needs a multibillion-dollar facility that can best be built through a fully fledged international collaboration. Other energy research has infrastructure requirements, too. But most of the parties to the ITER agreement are neglecting them, allowing energy research and development to drift along listlessly. In so doing, they betray an appalling lack of seriousness in confronting one of the century's most pressing problems. ■

"In Europe, energy research programmes have been eviscerated, despite political leaders' claims to be taking climate change seriously."



A little regulation

Regulators are beginning cautiously to navigate the uncharted waters of nanotechnology.

Scientists, businesses and politicians are growing increasingly interested in the safety of nanomaterials. Panel after panel has been convened to discuss whether substances engineered on the nanometre scale pose different environmental, health and safety risks from those engineered on larger scales. Report after report has called for greater scrutiny of any possible effects (see *Nature* **444**, 267–269; 2006).

So it should come as no surprise that US regulators have finally agreed on their first ruling concerning nanoparticles. In a decision that came to light last week, the Environmental Protection Agency (EPA) plans to regulate products marketed as antimicrobial because they contain nano-sized particles of silver. In essence, companies will no longer be able to sell things such as odour-eating silver shoe inserts without being able to prove — in a manner yet to be defined — that the particles will not harm the environment.

The ruling stems from a challenge over a Samsung washing machine, which is claimed to release silver ions to sterilize clothes. Environmentalists challenged an EPA decision last year that the washing machine was a 'device' and so did not need to be regulated. (Typically, devices that release pesticides are not regulated, but the pesticides are.) The agency reconsidered and now says the silver nanoparticles can be considered to be a pesticide, and so need to be regulated.

Companies that make and market nanoproducts have yet to see how environmental concerns may affect their business plans (see *Nature* **443**, 137; 2006). But the EPA is, in this case, taking a step in

the right direction. Clearly, many details remain to be worked out; the new ruling, for instance, leaves a significant loophole, as companies will be able to stop marketing their devices as antimicrobial and continue to sell them without regulation. But despite the criticism often hurled at them, regulators (at least in the United States and Europe) have a reasonably good track record of evolving their rules to protect the public interest.

The impact of all this on nanotechnology researchers and businesses will depend on how regulatory agencies cope with the dizzying array of products that will be produced under the wide rubric of 'nanotechnology'. If regulation is to be fair and effective, agencies will have to work closely together. Nanosilver, for instance, falls under the purview of the EPA when the particles could be released into the environment, such as in waste water from a washing machine. But it becomes the responsibility of the Food and Drug Administration for medical applications, such as when nano-sized silver particles are used on bandages to kill germs. Regulators from various agencies routinely attend nano-safety seminars and background briefings together; soon we will discover how much they have learned.

The scientific community, for its part, plays an important role in the development of technologies, including nanotechnology. But that doesn't mean it supports every application. Like everyone else, scientists want to see intelligent regulation. They should not interpret the EPA decision as negative: it is the first, necessary step along what promises to be a lengthy road. Regulators cannot begin to work their way through different health and safety risks until the first rules are on the table. Now they are. ■

"The Environmental Protection Agency's decision is the first, necessary step along what promises to be a lengthy road."

Not the end of an era

The Mars squadron loses its commodore.

The loss of the Mars Global Surveyor, a NASA spacecraft that has stopped talking to its controllers, is a sad event. It is impossible not to feel a sentimental 'brave little toaster' bond to a spacecraft that served its Earthly masters so faithfully for so long, and revealed so many wonders. And yet one cliché of loss often used in such situations is, on this occasion, happily inappropriate: this is most definitely not the 'end of an era'. Although the Surveyor is gone, almost ten years to the day after it was launched, the era it ushered in — that of the first continuous, if virtual, human presence at another planet — is still in full swing (see page 526).

For most of the 1980s and 1990s there was hardly any news from Mars. Since the Mars Global Surveyor started doing science in 1999 (getting into the correct orbit was a time-consuming procedure), the flow of data has been uninterrupted. Its cunningly designed camera revealed details of the surface never seen before, including fresh-looking gullies that appear to have been cut by water. Its magnetometer picked up the planet's intriguingly patterned magnetic field. Its laser

altimeter provided a topographic map far more accurate than anything that had come before. These instruments on their own would have revolutionized our view of the planet.

But they were not on their own. New spectroscopic capabilities were added by Mars Odyssey in 2002 and by the European Space Agency's Mars Express in 2004, which also provided a radar and a new camera. Information from the ground at two sites was provided by NASA's remarkably enduring Mars rovers, Spirit and Opportunity. This summer the heavyweight Mars Reconnaissance Orbiter joined the squadron with perhaps the most impressive instruments to date. The Phoenix lander is due to join them in a couple of years, and a more capable rover, Mars Science Laboratory, two years after that.

Mars Global Surveyor was the first of a sequence of 'faster, better, cheaper' NASA missions. This approach was subsequently renounced by the space agency, in large part because of the loss of the Surveyor's successors, Mars Climate Orbiter and Mars Polar Lander. But the idea of a continuous presence has persisted, with the various spacecraft calibrating and complementing each other's results, relaying each other's data when necessary, even taking pictures of each other. The fact that two spacecraft could help search for their fallen comrade is testament to the new age of exploration that the Surveyor inaugurated — even if, on this occasion, their efforts were in vain. ■

RESEARCH HIGHLIGHTS

DEVELOPMENTAL BIOLOGY

Secrets of the heart

Cell doi:10.1016/j.cell.2006.10.029 and 10.1016/j.cell.2006.10.028 (2006)

The heart's three major muscles can all grow from a single cell type, researchers have shown. The finding will help research into regenerative heart therapies.

Previously it was thought that each of the heart's muscle types came from distinct source cells. But, using genetic mapping techniques, Kenneth Chien of the Harvard Stem Cell Research Institute in Boston, Massachusetts, and his team identified a cardiac progenitor cell that can differentiate into cardiac, endothelial or smooth-muscle cells. These cells may be the parent of a second cell type — identified by Stuart Orkin of the Howard Hughes Medical Institute, Boston, and his colleagues in a separate paper — that differentiates into two of the heart muscle types.

CANCER BIOLOGY

Orient distress

J. Cell Biol. **175**, 547–554 (2006)

Many of the cells in our body use tiny hair-like antennae called monocilia as sensors. Now a team of researchers has shown how these monocilia go wrong in an inherited cancer syndrome called von Hippel–Lindau disease.

Thomas Benzing of the University Hospital Freiburg, Germany, and his colleagues examined how a faulty protein, pVHL, which underlies the disease, knocks out the formation of monocilia in kidney cells. The loss of these cilia contributes to the development of kidney cancer.

They found that pVHL localizes to the monocilia. There, it both interacts with other proteins involved in the cilia's formation and helps to orient microtubules, which form the scaffolding needed to build the protrusions.

Whole wheat

Science **314**, 1298–1301 (2006)

Crop researchers have discovered a gene in wild wheat that, when bred into cultivated varieties, boosts the grains' nutritional value.

Varieties of wheat (pictured) cultivated for making pasta and bread account for around 20% of all calories consumed worldwide. But these strains have a mutation that reduces their protein, zinc and iron content, report researchers led by Jorge Dubcovsky of the University of California, Davis.

The team identified the mutated gene as *NAM-B1*, which controls the redistribution of nutrients to the grains when the leaves die off. They also found that wild strains of wheat have an intact version of this gene, which can be crossed into domesticated strains to increase levels of micronutrients and proteins while retaining farming-friendly characteristics such as high yield.



D. REEDE/SPL

BIOMATERIALS

Copper-tinted roots

Biochemistry **45**, 14223–14231 (2006)

Marine mussels may have copper atoms to thank for their tenacious grip, say researchers at the University of California, Santa Barbara.

Mussels take hold of a surface by laying down a sticky plaque and then planting 'feet' into this layer. Hua Zhao and J. Herbert Waite studied the chemical structure of the plaque matrix protein mcfp-4 produced by the California mussel (*Mytilus californianus*), noting that it contains a motif rich in the amino acid histidine that is repeated about 36 times. Such units can bind copper atoms.

The researchers suggest that similar copper-binding units might exist in the mussel's 'foot' threads, creating multiple anchor points that bind its foot to the plaque.

GENETICS

Blue eyes or brown?

Am. J. Hum. Genet. (in the press)

Whether you have blue eyes or brown depends largely on three 'letters' in your genetic code, buried within a gene on your fifteenth chromosome. So conclude Richard Sturm of the University of Queensland in Brisbane, Australia, and his colleagues.

The researchers had previously shown that a particular region of the *OCA2* gene

could explain around three-quarters of the variation in human eye colour. They have now zoomed into this region, using data from 3,839 twins and their family members. They found three single-letter changes (known as single nucleotide polymorphisms, or SNPs) that are strongly associated with having blue eyes. The next challenge is to decipher the molecular mechanisms that translate the genetic differences into differences in iris pigment.

NANOTECHNOLOGY

Ultimate strength

Nano Lett. doi:10.1021/nl0619397 (2006)

Semiconducting nanowires have displayed their maximum theoretical strength, report John Boland of Trinity College Dublin, Ireland, and his colleagues.

The germanium nanowires had been grown by a 'supercritical fluid–liquid–solid mechanism'. The team used the tip of an atomic force microscope to bend tethered wires, which snapped without showing any substantial plastic deformation. This suggests that defects in the wires' crystal structure, which would weaken the material, played no role in their response to stress. Indeed, the wires' breaking point was close to the predicted ultimate strength for germanium nanowires; many other semiconducting nanowires snap at just



PHOTOLIBRARY.COM

15% of their theoretical limit.

The wires' elastic resilience could be useful in applications such as electromechanical oscillators, the researchers say.

CANCER BIOLOGY

Attractive prospect

Nature Cell Biol. doi:10.1038/ncb1507 (2006)

Researchers in Japan have discovered a mechanism whereby cancers produce long-distance signals that prepare the ground for a deadly tumour invasion of the lungs. They have also shown how these signals can be interrupted.

Yoshiro Maru of the Tokyo Women's Medical University and his colleagues found that primary tumours in mice secrete factors that increase the expression in the lungs of two chemoattractants. These proteins encourage the tumour cells to migrate into lung tissue.

Antibodies that bind the chemoattractants slashed the level of lung metastases in tumour-bearing mice by more than 80%. The researchers suggest that the chemoattractants, or the circuit of signals that induces them, provide a new clinical target for the prevention of cancer metastasis.

GENETICS

Forced apart

Science **314**, 1292-1295 (2006)

Fruitflies have provided the first example of a genetic mechanism that helps to drive a wedge between diverging species.

Hybrids of two different species are often born dead or sterile, enforcing the species' split. Their reduced viability is thought to result from interactions between gene pairs, called Dobzhansky-Muller genes, that have functionally diverged in the two species.

Daniel Barbash and his colleagues at Cornell University in Ithaca, New York, report the first known pair of Dobzhansky-Muller genes in the related fruitflies *Drosophila simulans* and *D. melanogaster*. They show that interactions between the protein products of the two genes *Lethal hybrid rescue* and *Hybrid male rescue* kill hybrid flies. As these proteins are known to bind to chromosomes, the researchers suspect that the lethal effect is due to changes in chromosome structure or shape.

NANOTECHNOLOGY

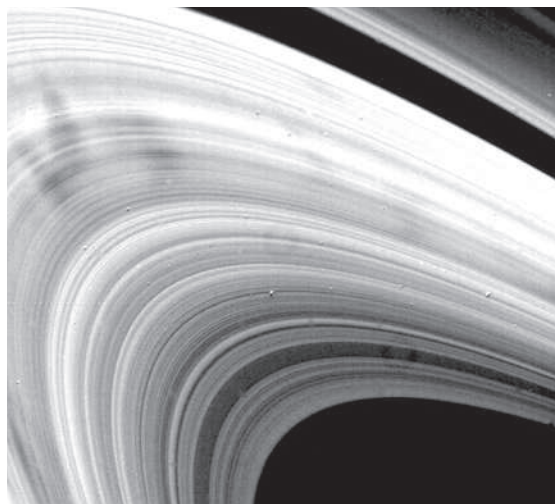
Grown from seed

J. Am. Chem. Soc. doi:10.1021/ja065767r (2006)

Just as gardeners can sow seeds to grow flowers of a particular colour, so chemists may soon be able to buy packets of seeds for growing specific varieties of carbon nanotube.

Nanotubes come in different widths and can vary in the amount of twist in their carbon walls. These characteristics determine the tube's properties — in particular, whether it is metallic or semiconducting. Researchers may want nanotubes of one type, but current synthesis methods give a mixture.

To solve this problem, James Tour and his co-workers at Rice University in Houston, Texas, divided a nanotube into short segments that can act as 'seeds' for growing identical tubes. They attached iron to the end of each segment, as this catalyses the growth of the seed when it is fed with a carbon-rich vapour. Cuttings can be taken from the resulting nanotube, from which further tubes can be grown if needed.



NASA

PLANETARY SCIENCE

Caught in the spokes

Geophys. Res. Lett. **33**, L21202 (2006)

The mysterious and transient 'spokes' that cross Saturn's rings (pictured above) might be caused by thunderstorms in the planet's atmosphere, researchers suggest.

Geraint Jones of the Max Planck Institute for Solar System Research in Katlenburg-Lindau, Germany, and his colleagues propose that electron beams discharged by thunder clouds are funnelled along the planet's magnetic-field lines and into the rings, where they blast out streams of dust.

Others have suggested that the dust streamers are created by meteorites ploughing through the rings, but Jones and his team contend that their theory can better explain why the spokes intensify over an hour or two. NASA's Cassini spacecraft has detected electron beams in orbit around Saturn, but there is as yet no direct evidence linking such beams either to thunderstorms or to spokes.

JOURNAL CLUB

Dolores R. Piperno

National Museum of Natural History, Washington DC, and Smithsonian Tropical Research Institute, Balboa, Republic of Panama

An archaeologist tells how her research interest sheds light on the history of her favourite fruit.

My mother gave me a lot of sage advice during my childhood. This sometimes countered prevailing wisdom, such as her recommendation that I should eat bananas, not apples, every day.

The nutritional qualities of bananas were underappreciated when I was growing up in Philadelphia, Pennsylvania, although their value as a crop had been recognized millennia before.

A recent paper (B. Lejju *et al.* *J. Archaeol. Sci.* **33**, 102-113; 2006) reports that bananas were being cultivated in Africa as long as 5,000 years ago.

The investigators reached this conclusion by studying phytoliths — highly durable pieces of silica that form in plant cells — that had been dug up in Uganda. I use phytoliths (and starch grains) in my own research into the history of agriculture, so it was more than just my fondness for bananas that drew me to this paper.

Phytoliths tell us which crops grew where and when, and so in turn can teach us about human cultures. Part of what makes this finding interesting is that bananas aren't native to the African continent. Wild bananas grow in eastern Asia, Australia and Melanesia. Hence the Ugandan bananas represent ancient human transport, probably through trade across the Indian Ocean with societies that were already growing the plant.

The age of the banana phytoliths also rivals presently available dates for the domestication of important, indigenous plants such as sorghum. However, I suspect that when phytolith and starch-grain analysis is applied to the history of native African crops, the date of their domestication will be pushed further back.

NEWS

Methane quashes green credentials of hydropower

At the time, it must have sounded like a sensible case of sustainable development. During the 1980s, about 2,500 square kilometres of Amazonian rainforest was flooded to create the Balbina dam to feed the energy demands of the Brazilian city of Manaus. A sizeable chunk of rainforest was lost, but Brazil gained access to a non-polluting energy source. It's a compromise Brazil has made many times; more than 80% of the country's domestic electricity is generated by hydropower plants.

Yet the clean, green image of dams may have been seriously overstated. Researchers are gathering in Paris next week to discuss greenhouse-gas emissions from tropical reservoirs.

Some of the latest findings point to a disturbing conclusion: that the global-warming impact of hydropower plants can often outweigh that of comparable fossil-fuel power stations. If that's correct, current energy strategies, particularly in developing nations, will need to be rethought.

The problem lies with the organic matter in the reservoir. Large amounts are trapped when land is flooded to create the dam, and more is flushed in after that. In the warm water of tropical dams, this matter decays to form methane and carbon dioxide. Although both

are greenhouse gases, the main worry is methane, which has more than 20 times the warming impact of carbon dioxide over a 100-year period. In the specific case of Balbina, there is now a rough consensus: in terms of avoiding greenhouse-gas emissions, a fossil-fuel plant would have been better.

But that is where the agreement ends. On one side of the debate is Philip Fearnside, a conservation biologist at the National Institute for Research in the Amazon in Manaus. His work, based mainly on theoretical calculations, looks at water leaving dams. Many dams release water from several metres below the surface, so the flow goes through an abrupt

pressure change. Fearnside calculates that this causes methane release, much as carbon dioxide fizzes out when carbonated drinks are opened. His latest results suggest that a typical tropical hydropower plant will, during the first ten years of its life, emit four times as much carbon as a comparable fossil-fuel station.

Lining up against him in a decade-long dispute are Luiz Pinguelli Rosa and his colleagues at the Federal University of Rio de Janeiro, who accuse Fearnside of exaggerating reservoir emissions. They complain in particular that

"If these estimates are correct, figures for annual global methane emissions need to be increased by a fifth."



Fearnside has extrapolated from measurements taken on the Petit Saut dam in French Guiana; the data were taken in the years immediately after the reservoir was created, when the store of organic matter would have been greatest.

With few data sets available on tropical dams, the debate has increased in acrimony without approaching a conclusion. Environmental groups question the impartiality of Rosa's work, which is funded in part by the hydropower industry. Rosa strongly denies any bias, and in turn accuses Fearnside of seeking to show that "something is wrong with dams".

The Paris meeting, which runs on 5–6 December and is organized by the United Nations

Preprint analysis quantifies scientific plagiarism

How often do researchers plagiarize each other's work? The question has previously been almost impossible to answer, as no large-scale survey of the practice had been conducted. But a computer scientist has now examined more than a quarter of a million documents from a physics preprint server. The results contain the comforting news that blatant deception is rare, but suggest that minor acts of misconduct may be more common than was previously thought.

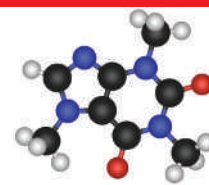
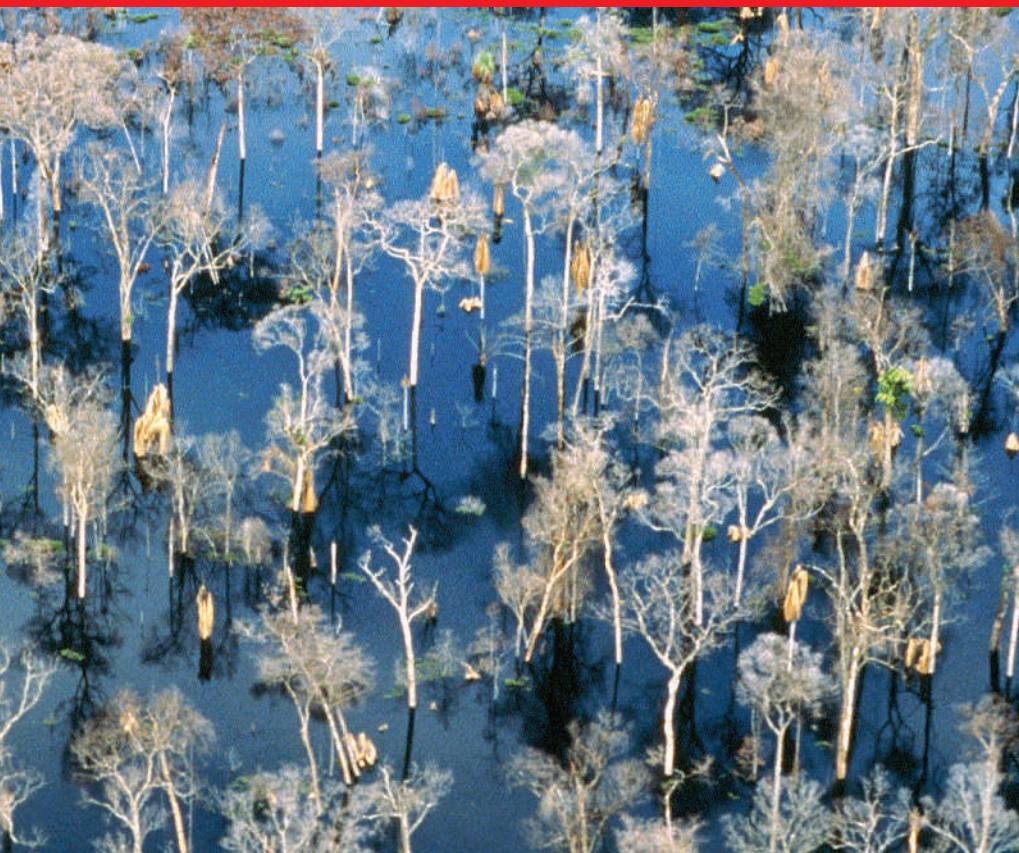
Student plagiarism can often be checked using specialist databases of essays available for sale online,

but plagiarism in published research is harder to police. Many publishers don't allow search engines to index the full text of their papers, so it's impossible to run electronic checks on new studies. Those small surveys that have been done revealed little evidence of plagiarism, but suggested that duplicate publications — in which a significant amount of an existing paper by the same author is reused without providing a reference to the original — could make up about 10% of the literature in some fields (see *Nature* 435, 258–259; 2005).

Firmer numbers can now be put on

those estimates, thanks to the work of Daria Sorokina, a PhD student at Cornell University in Ithaca, New York. Sorokina's software trawled more than 280,000 entries in arXiv, a database of mainly physics, maths and computer-science preprints maintained at Cornell. Her code divides documents into seven-word chunks and looks for pairs of papers that share a suspicious number of such chunks (common phrases such as "this work was supported in part by" are excluded). The result is a list of possible plagiarisms or, if the documents share a common author, duplicate publications.

The search turned up 677 examples of possible plagiarism, of which Sorokina and her colleagues took a close look at 20. Only four were innocent mistakes, such as articles that quoted text from a third scientist. Three of the others were judged to be 'serious plagiarism' in which one article was essentially a copy of another, and in the others, parts of the paper such as the introduction or related work sections had been copied without appropriate references being given. If the analysis scales up, then just 0.2% of arXiv documents contain plagiarism. The results will be presented at the



MATERIALS RESEARCH SOCIETY

Get the latest news from the Boston meeting.
<http://blogs.nature.com/news/blog>

global methane emissions (which do not generally include dam emissions) need to be increased by a fifth. Even extrapolating Rosa's figures gives Cullenward a total of 23 million tonnes.

Many think enough is known to start acting now. Some worry about the huge dam projects tentatively planned for tropical areas, such as a \$5-billion project on the Congo river. Another concern is the Clean Development Mechanism (CDM), a system that allows developed nations to fund clean-energy projects in developing nations in return for credits that can be used to meet Kyoto Protocol targets. Current rules allow certain hydropower projects to be funded under the CDM, a situation some scientists and environmental groups would like to see revised.

But matters are unlikely to change without more data, so researchers at the UNESCO meeting will discuss which questions to prioritize and how best to work together. More substantial progress could begin in 2008, when the Intergovernmental Panel on Climate Change (IPCC) will decide whether or not to start work on a special report on renewable energy. Previous IPCC special reports have had significant political impact, and the dams question is likely to fit very well into the scope of the proposed energy study, says Bert Metz, a climate-policy expert at the Netherlands Environmental Assessment Agency and co-chair of one of the IPCC's three working groups.

Jim Giles

1. Guérin, F. *et al. Geophys. Res. Lett.* **33**, L21407 (2006).
2. Kemezis, A., Forsberg, B. R. & Melack, J. M. in *Proc. 8th Int. Conf. Southern Hemisphere Meteorology and Oceanography*, Foz do Iguaçu, Brazil, 24–28 April 2006, 663–668 (INPE, São José dos Campos, Brazil, 2006).

The greenhouse-gas emissions from regions flooded by dams may have been grossly underestimated.

Educational, Scientific and Cultural Organization (UNESCO), is unlikely to settle their dispute, but researchers will discuss new methane data. On 14 November, for example, Frédéric Guérin of the Laboratory of Meteorology in Toulouse, France, and his colleagues published results on methane release from sites downstream of three tropical dams¹. They found that so much methane builds up in the dam that downstream emissions, which are rarely factored into estimates of a reservoir's impact, should account for between a tenth and a third of total emissions. Another new paper estimates

that, for Balbina, downstream emissions alone have the same greenhouse warming potential as 6% of all the fossil fuels consumed by São Paulo, a city of more than 11 million people².

Even without these downstream emissions, the global impact of dams may be significant. Danny Cullenward, an energy-policy expert at Stanford University, has made preliminary calculations of the impact of Fearnside's findings. Cullenward stresses that more data are needed, but his estimates suggest that dams release between 95 million and 122 million tonnes of methane per year. If correct, estimates of annual

IEEE International Conference on Data Mining, to be held on 18–22 December in Hong Kong.

Results for duplicates are potentially more alarming but harder to assess. Sorokina identified 30,316 pairs where one was largely a copy of the other — more than 10% of the database. But arXiv differs from journals in that researchers submit conference proceedings as well as the journal papers that are derived from them. Paul Ginsparg, a Cornell physicist who worked with Sorokina on the survey, says the “vast majority” of duplicates found are of this type, but adds that he was surprised at the number of student theses that included material copied verbatim from other sources.

Despite the leap forward provided by the arXiv survey, many issues remain unresolved. The survey picks up only cases where the source that has been plagiarized

“This may be the most accurate global estimate that we have for plagiarism in the scientific literature.”

is also present on the arXiv database (although in many fields arXiv has near-complete coverage). And the software is unable to pick up ‘intelligent plagiarism’, where material copied from another author is reworded.

Researchers may also behave

differently when submitting to arXiv compared with peer-reviewed journals, and different rates may exist for biologists, who rarely use preprint servers. Plagiarism in biology could in principle be studied using PubMed Central, an archive of journal papers maintained by the US National Institutes of Health. Ginsparg says he has discussed this with PubMed's staff, but such a survey would currently be of limited use, as only a small fraction of papers are placed in the database after publication.

Larry Claxton, a toxicologist at the US Environmental Protection Agency in North Carolina, who has studied plagiarism in the life sciences, says he would like to see a more detailed examination of the duplicates in the arXiv study before

interpreting the results. But despite the limitations, he says, “this may be the most accurate global estimate that we have for plagiarism in the scientific literature”.

Ginsparg says he would now like arXiv to scan all new entries using Sorokina's software and alert the author in the case of suspicious overlap. The researcher would then have the option of rewriting the paper or, if they felt the overlap was justified, submitting as usual. Ginsparg says the system could be in place by the middle of next year. Preliminary discussions have also taken place with CrossRef, a publishers' group, about whether journals could work together to implement a similar scheme.

Jim Giles

ON THE RECORD

“Wouldn't it be fascinating to take a Fourier transform of those waves?”

The words with which science blogger Sean Carroll, of Cosmic Variance, captured the heart of Jennifer Ouellette, who blogs at Cocktail Party Physics, while watching a Pacific sunset. They are to marry next year.

“This is a guy who takes a neuropeptide and a prairie vole and spins from them science fiction.”

Journalist Amanda Schaffer is outraged that Eric Keroack has been chosen to head the US federal family-planning programmes. Keroack says that research in voles has led him to believe that promiscuity makes people less able to bond with their children.

“For the time-being our immediate solution is to send them to the taxidermists.”

Muhedin Abdulaziz, of Addis Ababa's zoo, on the problem of having more Abyssinian lion cubs than the zoo can handle.

SCORECARD

Amateur fusion

 Schoolboy Thiago Olson of Oakland Township, Michigan, has reportedly become the 18th amateur scientist to achieve a low level of nuclear fusion using a home-made set-up.

NUMBER CRUNCH

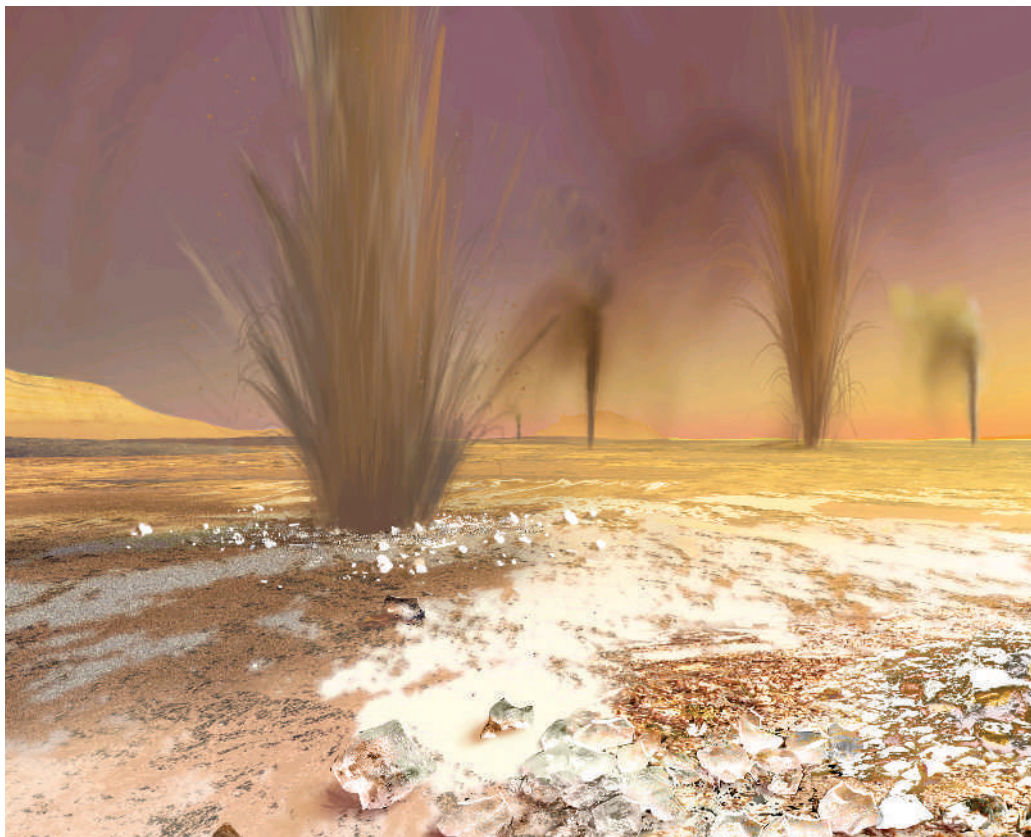
US\$20 million is the ticket price for a trip to the International Space Station on a Russian spacecraft.

2,080 is the number of years a Russian cosmonaut would have to bank his or her entire \$810-a-month salary to save that much money.

0 is the number of cosmonauts recruited from Russia's top universities through outreach programmes this year.

Sources: Slate, Detroit Free Press, The Times.

NASA/R. MILLER



Spouting forth: artist's impression of dusty gas jets found by Mars Global Surveyor at Mars's south pole.

A space dilemma: extend missions or start afresh?

Scientists said a sad goodbye last week to the orbiting spacecraft Mars Global Surveyor, which had gone missing two weeks earlier.

The mission has been portrayed as heroically long-lived, surpassing all expectations and being extended four times until scientists finally lost contact with it just before the tenth anniversary of its launch. It's a familiar story for NASA missions — the Mars rovers Spirit and Opportunity have both lasted ten times longer than their 90-day intended lifetimes and continue to roam the planet, and the agency is spending billions of dollars on a shuttle mission to keep the Hubble Space Telescope running against all the odds.

But however attached we might get to old missions, does it really make financial sense to run them until they break, get lost or fall into a hole too big to get out of?

Extending missions is a trade-off between wringing as much science as possible out of a

craft and redirecting resources into new missions. And at NASA, keeping missions going is seen as well worth the money. An extended mission typically costs a few million dollars a year to run, whereas a new launch can cost hundreds of millions or more. Critics might argue that the most important results from the Mars rovers, such as clinching the case that liquid water once existed on the planet, came from the early part of the mission. But Bruce Banerdt, rover project scientist at the Jet Propulsion Laboratory in Pasadena, California, claims that every time the rovers move a few hundred metres, it's like landing a new lander: "It's scientifically cost-effective."

James Green, director of NASA's planetary-science division, acknowledges that with an extended mission you are restricted to the science the craft was designed for, rather than choosing what needs to be done. But Green insists that, because they are so much cheaper than a



SPACE-AGE RELAY

NASA lost contact with Mars Global Surveyor on 2 November. Despite bombarding the craft with more than 800 command files, the last scientists heard was a possible faint signal on 5 November.

Mars Global Surveyor made the first reliable topographic map of the whole planet, and revealed swathes of the surface in great detail. It also allowed the first analysis of Mars's ancient magnetic fields. But perhaps its most dramatic discovery was of fresh gullies in the walls of some craters and on other slopes that seemed to have been formed by fluid flow.

More recently, the craft was helping Mars Reconnaissance Orbiter, which arrived in March, to scout out a landing spot for the next generation of Mars probes: Phoenix, due to arrive in 2008, and Mars Science Lab, which will touch down in 2010.

Mars Reconnaissance Orbiter's mission is scheduled to run until 2010, but is expected to last far longer. The planet is currently circled by two other craft, NASA's Mars Odyssey and the European Space Agency's Mars Express. The overlap between Mars Global Surveyor and these missions marks the first time that one craft orbiting a planet (other than Earth) has passed the baton to another. **K.S.**

new mission, extended missions are worth the effort. NASA's mission scientists bid for extensions once the primary science objectives have been fulfilled, and missions are judged on their scientific impact to date: "science per dollar", he says.

Possible extensions are not mentioned in the initial mission planning stages, so designing missions to last longer than planned is an easy way for mission scientists to buy into future budgets. There's also the public-relations value — in terms of public image, it is better to say a mission will last for five years and then extend its life, than to say it will last for ten years and lose it after nine. The latter scenario would be seen as a failure, says Lori Garver, a former NASA associate administrator. She sees no problem with deliberately underestimating the life of a spacecraft. By the time a mission is ready to be extended, the big pots of money have already been spent on it, she says.

Attitudes at the European Space Agency (ESA), which has a smaller budget and flies fewer missions (17 spacecraft are currently operational), are slightly different. Letting missions run until they fail, with no predefined notion of how long that might take, is not necessarily a sustainable strategy, says Giovanni

Bignami, chairman of ESA's space-science advisory commission. "By continuing to extend missions you eat into resources that could be used to do something new," he explains.

Up to 10% of ESA's budget is spent on keeping old missions alive, and Bignami says that no missions have been ended while they were still operational. But the agency expects that with several new craft due to launch in the next few years — including BepiColombo bound for Mercury, and Gaia, which will create a three-dimensional map of our Galaxy — it will become more difficult to keep old missions running.

ESA approves missions for two years at a time, and some are definitely worth extending, says Bignami; for example, the XMM-Newton X-ray satellite, which is looking at black holes and exploding stars. He reckons that other missions might be easier to finish, such as Venus Express — there's a limit to how much useful data can be gathered by continuing to orbit Venus with the same instruments.

Although many are sad to see Mars Global Surveyor go, Bignami also anticipates some happy faces at NASA, as an old mission is finally abandoned and new ones can begin. ■

Katharine Sanderson

See Editorial, page 520.

"Missions are judged on their scientific impact to date — science per dollar."

Amazon puts network power online

GETTY IMAGES



Plug in: industry supercomputer power on the desktop PC could have a big impact on scientific research.

When Dutch computer scientist Rudi Cilibrasi needed hundreds of hours-worth of computing time to test a data-mining algorithm earlier this month, he went not to his IT department but to Amazon.com. He paid \$60 with his credit card, and in minutes had the equivalent of ten servers installed, which crunched through his job in a couple of days — ten times faster than his desktop PC would have managed.

Large web companies such as Google, eBay and Amazon have far more computing power at their disposal than any academic networks, and have become leaders in massive-scale distributed computing. Many of their innovations can help scientists, and Amazon's computing-on-demand service, which has been running since August, is no exception. It enables customers to create multiple virtual computers on Amazon's massive computing infrastructure for \$0.10 per computing hour, and to store data for \$0.15 per gigabyte per month.

The service is still in a test phase, so few scientists have even heard of it yet, let alone tried it. But it is a movement that experts believe could revolutionize how researchers use computers. In future, they will export computing jobs to industry networks rather than trying to run them in-house, says Alberto Pace, head of Internet services at CERN, the European particle-physics laboratory near Geneva. CERN has built the world's largest scientific computing grid, bringing together 10,000 computers

in 31 countries to handle the 1.5 gigabytes of data that its new accelerator, the Large Hadron Collider, will churn out every second once it is switched on next year.

"I see no reason why the Amazon service wouldn't take off," Pace says. "For a lab that wants to go fast and cheaply, this is a huge advantage over buying material and hiring IT staff. You spend a few dollars, you have a computer farm and you get results."

Cilibrasi, a researcher at the National Institute for Mathematics and Computer Science in Amsterdam, was using Amazon's service to test an algorithm aimed at predicting how much someone will like a movie based on their current preferences. He says he is a convert: "It's substantially more reliable, cheaper and easier to use [than academic computing networks]. It opens up powerful computing-on-demand to the masses."

It's not just computer scientists who could use cheap computing power. Climate researchers can run global-warming models, cosmologists might simulate the inside of a supernova or biologists could scale up the workings of a liver cell to the whole organ.

The South African National Bioinformatics Institute at the University of Western Cape, Belleville, has already been testing Amazon's system to power large-scale genome comparisons. The pay-as-you-go system offers com-

puting power and bandwidth that the institute could not afford to maintain itself, says Inus Scheepers, a systems administrator there.

Cost is certainly one reason observers are excited about Amazon's system. Other companies, including Sun Microsystems, offer computing-on-demand, but Amazon's service costs a tenth as much. But the main attraction is Amazon's use of 'virtualization' technologies, which many predict will change not just research but computing itself.

Virtualization uses a layer of software to allow multiple operating systems to run together. This means that different computers can be recreated on the same machine. So one machine can host say ten 'virtual' computers, each with a different operating system.

That's a big deal. Running multiple virtual computers on a single server uses available resources much more efficiently. But it also means that instead of having to physically install a machine with a particular operating system, a virtual version can be created in seconds. Such virtual computers can be copied just like a file, and will run on any machine irrespective of the hardware it is using. "In the past, we had to install hardware and software for each machine," says Pace.

Virtualization software is becoming widely available from companies such as Microsoft and VMware. Amazon uses Xen, an open-source system developed at the University of Cambridge, UK, which is fast becoming popular with researchers. As well as using Xen to

"You spend a few dollars, you have a computer farm and you get results."

enable virtual computers to run across a grid or cluster, despite different operating systems on the individual machines, researchers can also use applications developed on their lab computer.

At present, to run an application on a large scale, they often need to rewrite it. With virtualization, researchers can create a copy of their own machine and use it to run large-scale simulations or searches, and it should work exactly as it does in the lab.

Amazon's service combines the joys of virtualization with the huge computing power it has at its disposal. It's an approach that looks set to catch on. CERN has started an internal service similar to Amazon's, in which users can create or delete virtual machines on the fly. "Virtualization is revolutionary," says Pace. "It's clear that this is one way to do scientific research in the future."

Declan Butler



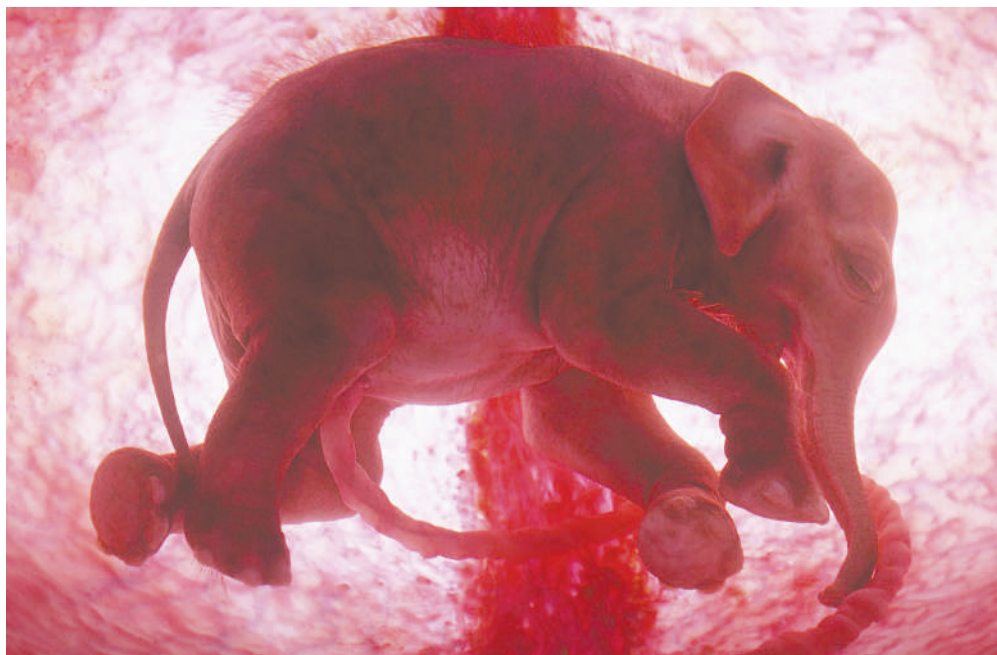
**HIV TO BE TOP HEALTH
ISSUE WITHIN 25 YEARS**
AIDS set to become world's
biggest problem disease.
www.nature.com/news

SNAPSHOT A womb with a view

Opening up an unseen world, this detailed ultrasound scan captures an elephant not in the wild — but in the womb. The image, which has been enhanced with computer graphics, shows a 75-kilogram elephant fetus already decked out with footpads and a trunk.

To get close enough to the fetus to generate images, ultrasonographers donned shoulder-length gloves and gave the pregnant mother an enema — before inserting an ultrasound probe up the length of her rectum. This fetus has been gestating for nearly 19 months and would have started kicking and exercising its legs about 14 months earlier. It still has three months before it is to be born.

The picture was captured as part of a documentary called *In The Womb*, in which ultrasound images of dolphins and puppies as well as elephants were pieced together into short films, culminating in footage of their births. The show is to be broadcast next month on the National Geographic Channel in the United States and Europe, and on Channel 4 in Britain.



The documentary also touches on the subject of evolution. At four months old, the elephant fetus temporarily develops kidney ducts normally found only in freshwater fish and frogs, hinting at its ancestors' aquatic origins.

Narelle Towie

NATIONAL GEOGRAPHIC

Past drought hints at Africa's future

An ecological survey of one of East Africa's worst-ever droughts provides a fresh look at how ecosystems respond to sudden, dramatic disturbances. The study documents a 19-year period in the late nineteenth century known to the native Masai people as *Emutai* — meaning 'to wipe out' — and it could help in determining how the landscape may respond to future climate shifts, as are predicted as a result of global warming.

In 1891, when the Austrian explorer Oscar Baumann travelled in Masailand in modern-day Kenya, he was shocked to see "women wasted to skeletons from whose eyes the madness of starvation glared". Drought had first struck in 1883; the rains later failed completely in 1897 and 1898; rinderpest and bovine pneumonia destroyed herds. The Masai, who traditionally rely on animals for all of their food, were faced with famine and smallpox.

The latest research, published in the *African Journal of Ecology* (44, 458–467; 2006) reveals the ecological after-effects. "It was already in the history books — I was trying to calibrate the palaeoecological record against historical records," says Lindsey Gillson of the University of Cape

Town, who carried out the study at the University of Oxford's Environmental Change Unit.

She travelled to Kenya's Tsavo National Park to collect sediment, pollen and charcoal samples to determine the extent of the drought. The abundance of pollen from marginal wetland plants shows that lake shores were retreating during the event, and the increased presence of charcoal shows that spontaneous fires were common — all features of a major drought.

Gillson's results confirm that the *Emutai* was indeed a 'large infrequent disturbance' — the kind of climatic event, such as droughts, floods or hurricanes, that can leave a lasting impression on the land. And it's the kind of event that experts predict will become more common as global warming proceeds.

The work hints at the effects that can be expected if the region is hit by another drought of this severity. "What we need is a kind of probability — if we got a drought like this, will x, y and z happen?" says Gillson. And now that she has identified the ecological signature of the

drought, looking back through the palaeoecological record should help to show how often such extreme events are occurring and whether that frequency is changing over time.

But it will be crucial, Gillson adds, to link the ecological data more firmly with climate predictions, so that local people can be prepared and adapt to these changes.

Helping Africans adapt to climate change is a priority, says Jon Lovett, an environmental researcher at the University of York, UK. "Given the close relationship between African livelihoods and ecology, African ecologists need to

consider urgently how to apply their skills and knowledge to the impact on humans of climate change," he says.

That is especially true for the Masai. Development agencies are now urging them to diversify from their reliance on livestock and embrace staple crops. With three million Kenyans suffering from the current drought, the calls for a change of habits are growing louder.

Michael Hopkin

**"African ecologists
need to consider how
to apply their skills
and knowledge to the
impact on humans of
climate change."**

Japan speeds up nuclear physics

No particle accelerator in the world is strong enough to create a usable beam of uranium ions. But that will change next month, when Japan switches on a huge facility of connected accelerators, to produce the world's most powerful beams of heavy radioactive isotopes.

Radioisotopes are forms of elements that are unstable because they contain either more or fewer neutrons than usual, and so undergo radioactive decay. Nuclear physicists are studying rare short-lived isotopes to understand their properties and how they are formed. The RIKEN research institute in Saitama, Japan, already has accelerators that can create the world's strongest radioisotope beams, but even these are only powerful enough to produce usable beams for the lighter elements.

But next month, RIKEN will switch on a major upgrade. The ¥44-billion (US\$378 million) Radioactive Isotope Beam Factory will add two more ring cyclotrons and the world's first superconducting ring cyclotron to the

existing linear accelerator and ring cyclotron. It will then be able to accelerate beams of any element up to uranium at 70% of the speed of light. The accelerated beams are smashed into a target such as beryllium to knock out neutrons and protons and create the desired radioisotopes.

The facility should open a new realm of astrophysics. "With this new facility, scientists at RIKEN have the opportunity to study nuclear isotopes that exist only in the hottest stars of the Universe," says John Schiffer, a senior scientist at the Argonne National Laboratory in Chicago, Illinois.

As well as exploring the formation of uranium, RIKEN plans to measure the properties of various very short-lived nuclei, as well as looking for 'magic numbers' of neutrons and protons that allow heavy nuclei to be surprisingly stable. These experiments will start from next year, with full operation scheduled for

2011. The facility makes Japan the world leader in the field, says Ysushige Yano, director of the RIKEN Nishina centre for accelerator-based science, adding that Japan's other big physics facilities have just been upgrades of US and

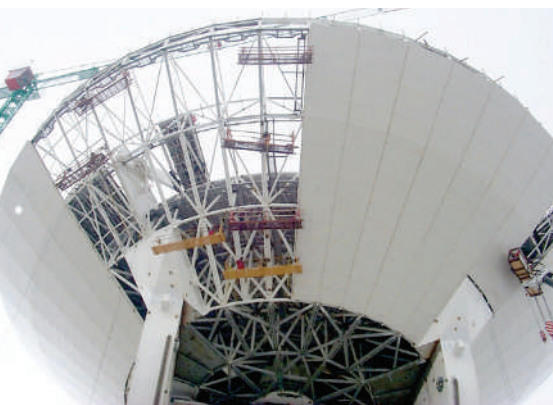
European versions. "But this time it is different," he boasts. "This time, Japanese scientists are leading the way."

Rivals aim not to let Japan savour its victory for long.

A US plan for a superconducting linear accelerator called the Rare Isotope Accelerator has stalled, at a proposed cost of \$1 billion. But France is expected to complete construction of its new radioisotope facilities, including experiments, by around 2012 and Germany by 2014. "In five or six years, Japan may lose the number one position," says Sydney Gales, director of the French heavy-ion accelerator GANIL in Caen.

Ichiko Fuyuno

"Scientists will be able to study nuclear isotopes that exist only in the hottest stars."



Eye on the sky: the Large Millimeter Telescope will study a range of cosmic phenomena.

Mexican telescope gets the president's blessing

Mexico's outgoing president Vicente Fox has inaugurated the Large Millimeter Telescope, a 50-metre dish located atop a dormant volcano in the state of Puebla.

The telescope will gather millimetre-wavelength light to study everything from comets to the birth of stars and the cosmic microwave background. Speaking at the ceremony on 22 November, President Fox told the crowd that the telescope would "put Mexico in the scientific and investigative vanguard in this field".

The US\$120-million telescope is a joint venture between Mexico's National Institute of Astrophysics, Optics and Electronics and the University of Massachusetts at Amherst. Construction began in 1997. Project scientists hope the commissioning phase will end in 2008, and that the telescope will see first light soon afterwards.

Online tool for Linux software proves popular

Make scientists' lives easier and you can suddenly find yourself in great demand. That's what Urban Anjar, an information-technology systems administrator at Kalmar University in Sweden, discovered this month.

Frustrated with repeatedly telling scientists where they could find software from various scientific fields for use with the Linux operating system, Anjar wrote a graphical interface that lets users automatically install the most widely used packages at the click of a mouse for one of the most popular versions of Linux, Ubuntu.

Anjar put his creation, Scibuntu, on the Internet on 3 November. "Then the craziness began," says Anjar. "I began to get mail from scientists, students and developers all over the world." A search on Google turns up almost 100,000 pages

that mention Scibuntu. Anjar has now transferred the project to the open-source repository Sourceforge, and hopes to enlist other developers to further improve his software.

► <http://sourceforge.net/projects/scibuntu>

Experimental journal puts science in the movies

Molecular biology is riddled with intricate protocols, and fine details can make or break an experiment. Now researchers can get every last bit of information for certain procedures at the newly launched *Journal of Visualized Experiments*. The online journal, which started publishing submissions this month, consists entirely of videos of scientists demonstrating basic biological procedures.

Some journals already allow scientists to submit videos to supplement the methods sections of their papers, but the new journal is the first to focus exclusively on this. It is the pet project of Moshe Pritsker, a postdoctoral researcher at Massachusetts General Hospital in Boston.

Pritsker hopes the journal will help scientists improve the reproducibility of their work, while also providing a window for the public on what researchers really do in the lab.

► www.myjove.com

Max Planck Society plans expansion to United States

Germany's flagship research organization, the Max Planck Society, is planning to open an institute in Palm Beach County in

Florida, having been courted by the state's governor Jeb Bush.

The Max Planck Society already has institutes elsewhere in Europe, but the proposed centre for biomolecular imaging at Florida Atlantic University would be its first on US soil. The centre would collaborate with an offshoot of the California-based Scripps Research Institute, which has already set up shop on the university's Jupiter campus (see *Nature* 442, 729; 2006).

Peter Gruss, the president of the Max Planck Society, announced the plan on 24 November to officials, who will vote on the proposal next year.

Slovenians engineer win in genetics competition

Cells engineered to intercept the immune system's excessive response to bacterial infection seized the grand prize in this year's international Genetically Engineered Machine (iGEM) competition.

The competition, now in its third year, challenges students to create biological devices using an arsenal of components known as BioBricks (see *Nature* 438, 417–418; 2006). These are genetic components with well-defined functions that can be assembled, like electronic parts on a circuit board, to perform more complex actions. The competition's aim is to foster the development of new tools for use in synthetic biology.

Students from the University of Ljubljana in Slovenia beat 36 other teams to take the top prize earlier this month. Other entries included *Escherichia coli* with the whiff of mint or banana, and a biological oscillator.

Chinese plant to give solar power a bright future

China is to construct one of the world's largest solar-power plants, according to a report from the state news source Xinhua on 21 November.

The 100-megawatt plant, costing 6 billion yuan (US\$760 million), will be built in the city of Dunhuang in northwestern China. It is expected to take five years to complete. Other countries are planning facilities on a similar scale, with Australia last month announcing funding for a 154-megawatt facility.

In January, China implemented a law aiming to make renewable energy account for 15% of the country's total energy consumption by 2020. But even this massive project is dwarfed by China's drive towards using fossil fuels. The nation is opening about 100 coal-powered plants every year, each of which will produce about 600 megawatts of power.



W. SCHROLL/CORBIS

BUSINESS

The bitterest pill

Drug companies lose hundreds of millions of dollars when large-scale human clinical trials fail. **Helen Pearson** examines whether alternative procedures could help avoid such disappointments.

Cape Town in October has many attractions: balmy temperatures, enticing beaches, ample wines and the indomitable presence of Table Mountain. But for the many neurologists attending the Joint World Congress on Stroke there late last month, it was a glum place to be.

The neurologists learned that a highly promising stroke drug called NXY-059 had dismally failed in a phase III clinical trial. The results, announced by AstraZeneca on 26 October, were particularly disappointing because they contradicted those of a previous phase III trial, which had shown that the same drug could help patients by quenching the free radicals that damage neurons after a stroke.

AstraZeneca's shares fell by 7.5% to \$61.38 on the day of the announcement. Shares in Renovis, the San Francisco biotechnology company that licensed the drug to AstraZeneca, plunged by more than 75%. Neurologists at the meeting were just as deflated. "We were all pretty down and asking what the hell's going on here," recalls Marc Fisher of the University of Massachusetts Medical School in Worcester.

The failed phase III trial shouldn't have come as too much of a surprise. More than 40% of drugs that enter phase III trials — the final and most extensive stage of clinical testing — are abandoned because they prove ineffective or unsafe (see graph).

Costing failure

Phase III trials are the priciest part of clinical development, making up an estimated 70% of a compound's clinical development costs. Improving their success rate would be "the single biggest factor in bringing down the cost of drug development", says Thomas Roberts, a senior healthcare analyst at the investment fund Noonday Asset Management in Charlotte, North Carolina.

Drugs with a novel mechanism of action are more likely to fail than those with an established way of working. Compounds to treat cancer and central nervous system disorders, including stroke, have particularly high failure rates — partly because animal models are such a poor mimic of the way these diseases affect humans.

Researchers have long sought a neuro-protectant drug that could salvage some of the brain tissues starved of oxygen after a stroke.

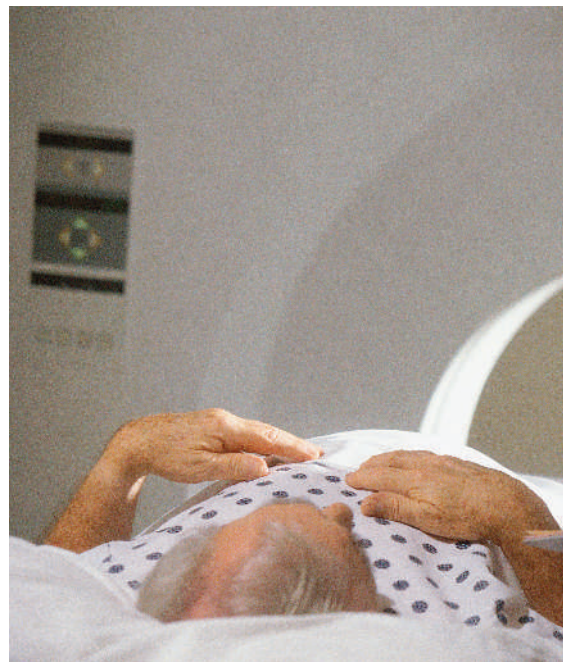
But of more than 100 drugs that have entered major clinical trials over the past couple of decades after strong results in animals, none has proven effective in patients. That's probably because the artery-clogging clots in the human brain are caused by a diverse collection of underlying diseases that are not reflected in animal models.

Even so, hopes were riding high for NXY-059, which had met particularly stringent tests in animals. That optimism seemed to be borne out after the first phase III trial in 1,700 patients (see K. R. Lees *et al.* *N. Engl. J. Med.* **354**, 588–600; 2006).

But there were warning signs from the first trial. It showed that the drug was statistically better than a placebo using one scale to measure improvement in a patient's disability — but it didn't significantly improve symptoms, using a second, separate, measure of its impact. This made at least some specialists question whether the drug would really improve patients' quality of life. AstraZeneca was "pushing the data very hard", says Peter Rothwell, who directs the Stroke Prevention Research Unit at the University of Oxford, UK — adding that, for him, "it wasn't convincing".

The second phase III trial, involving more than 3,200 people, showed that the drug was no better than placebo. So Rothwell and other experts now think that the earlier result was simply a statistical fluke. Like any result that is significant based on a statistical test, there is a small probability (about 4% in this case) that it seemed to help patients by chance even though the drug was actually ineffective.

But what of the phase III trials that fail at the first try? An April 2006 analysis of more than 1,600 phase III trials by management



Poor performance: about half of large-scale drug trials in humans fail to reach the approval stage.

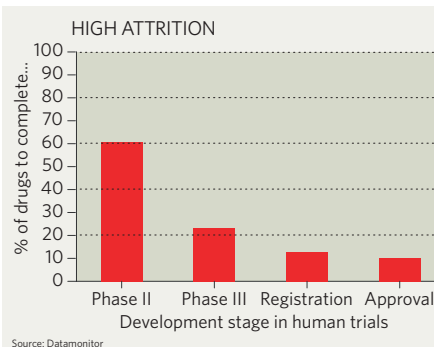
consultancy McKinsey found that 42% of them were unsuccessful. In half of these failures, the drugs were no more effective than a placebo; 30% raised safety concerns, and the remaining 20% were found to be no safer or more effective than existing drugs.

Proceed with caution

Some specialists argue that the failure rate could be cut by even more extensive and rigorous preliminary testing of drug candidates, in both animals and humans. To this end, a great deal of research effort goes into trying to improve animal models and, of course, to fully understanding a drug's mechanism to prove that it affects a molecule or process important to a disease. The McKinsey report also suggests that companies should act more quickly to identify high-risk compounds, and subject them to more stringent preliminary tests.

Ismail Kola, senior vice-president of basic research at Merck Research Laboratories in Rahway, New Jersey, says that pharmaceutical companies could decide to reject drugs at an earlier stage by carrying out quick 'proof of concept' studies in just a few patients without the expense of a large trial. He also advocates the development of more objective ways to gauge symptoms in patients, such as biomarkers in the blood, rather than inherently subjective ones, such as symptom scales. Kola says that drug companies have started to introduce these approaches over the past couple of years.

Even with more data on a drug, any decision





to withdraw it from testing can be clouded by wishful thinking. In biotechnology companies with only one or two drugs in development, the whole future of the company and its employees can hang on the success of a single compound; employees financially and emotionally invested in an experimental drug may ignore warning signs. In larger drug companies, too, those who have spent years working on a compound are obviously keen to see it succeed. "People get quite wedded to their products," Roberts says.

Another way to improve trial success rates, Roberts says, is to identify subgroups of patients in which a drug is more likely to work, using technology such as genetic tests or biomarkers. Some companies are reluctant to do this, fearing that it will reduce the market for which a drug can ultimately be sold. But Roberts has calculated in an economic model that this loss would be offset by money saved from otherwise unsuccessful phase III trials. "In my view it's one of the most powerful ways we could improve the situation," he says.

Such an approach offers a sliver of hope to those in the stroke field. After the poor phase III results, AstraZeneca cancelled the programme and it has no other stroke drugs in development. Few major pharmaceutical companies are pursuing neuroprotectants, although some academic researchers and smaller companies are. But researchers are still hopeful of finding subgroups that respond better to other experimental drugs — perhaps by identifying those best suited to treatment with brain imaging. Ultimately, the field must "deal with this wreckage" from the failure, says Fisher, and move on to new candidates. ■

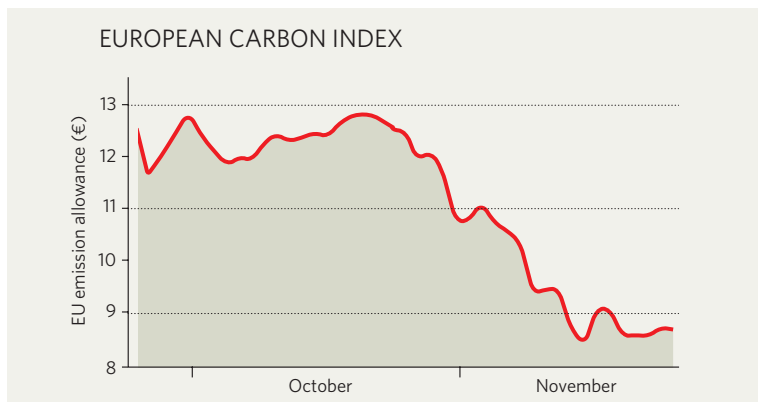
IN BRIEF

GENOMICS GRAB Illumina, the gene-chip maker based in San Diego, says that it will pay \$600 million to acquire Solexa, a manufacturer of gene-sequencing machines. The merger, expected to be complete by 31 March next year, will create the first company with technologies in large-scale genotyping, gene expression and gene sequencing — the main tools for exploiting information from the human genome. The move caps a year of significant gains for Illumina (see *Nature* 444, 256–257; 2006), which says it will retain Solexa's current operations in Hayward, California, and Cambridge, UK.

SUPERCONTRACTING The US Defense Advanced Research Projects Agency (DARPA) says it is granting almost \$500 million to IBM and Cray to press ahead with the development of supercomputers. The contracts place strong emphasis on making sure that it will be easy to write applications software that will actually make use of computers' vast processing power. DARPA says the computers will deliver performance of more than a petaflop, or 10^{15} floating-point operations per second.

MERGER MANIA Mergers and acquisitions of biotechnology companies soared in the second half of 2006, according to a report by Sanford C. Bernstein, an investment research and management firm based in New York. Since July, biotech firms have merged with — or, more frequently, been acquired by — other biotech or drug companies in nine deals worth \$24.3 billion. That compares with \$15 billion in similar deals in the whole of 2005, and \$2.9 billion in the first half of this calendar year. "We're in record territory in terms of both numbers and value of these mergers," says report author Geoffrey Porges.

MARKET WATCH



The price of European Union (EU) allowances for carbon dioxide emissions has steadily declined over the past six weeks, reaching an all-year low of less than €8.50 (US\$11.11) last week.

Even optimists are now having doubts about whether the market will recover again during the rest of the first phase of the EU's emission trading system, which finishes at the end of 2007. Most companies participating in the system own enough allowances for next year, and large energy producers have lost interest in backing the market by buying more allowances than necessary. As a result, analysts say, prices may go further downhill.

"Given the surpluses, the price should in theory approach zero," says Stefan Kleeberg, who watches energy markets for 3C Climate Change Consulting in Frankfurt, Germany.

The poor outcome of the United Nations' climate summit in Nairobi has not exactly stirred euphoria among emissions traders. Nonetheless, analysts reject the idea that the trading system might collapse altogether. "The books are closed for 2006 and 2007," says Kleeberg. "But questioning the efficiency of the system as such would be wrong."

Emissions allowances for the second trading period, from 2008 to 2012, are currently traded at around €17, meaning that the difference between phase I and phase II prices is at its largest ever. Reports last week that the European Commission will reject Germany's national allocation plan for 2008–12 — effectively demanding tighter caps on Europe's largest emitter — fuelled demand for phase-II allowances at European carbon exchanges.

Quirin Schiermeier

EUROPEAN ENERGY EXCHANGE

IN SEARCH OF LOST TIME

The ancient Antikythera Mechanism doesn't just challenge our assumptions about technology transfer over the ages — it gives us fresh insights into history itself. **Jo Marchant** reports.

It looks like something from another world — nothing like the classical statues and vases that fill the rest of the echoing hall. Three flat pieces of what looks like green, flaky pastry are supported in perspex cradles. Within each fragment, layers of something that was once metal have been squashed together, and are now covered in calcareous accretions and various corrosions, from the whitish tin oxide to the dark bluish green of copper chloride. This thing spent 2,000 years at the bottom of the sea before making it to the National Archaeological Museum in Athens, and it shows.

But it is the details that take my breath away. Beneath the powdery deposits, tiny cramped writing is visible along with a spiral scale; there are traces of gear-wheels edged with jagged teeth. Next to the fragments an X-ray shows some of the object's internal workings. It looks just like the inside of a wristwatch.

This is the Antikythera Mechanism. These fragments contain at least 30 interlocking gear-wheels, along with copious astronomical inscriptions. Before its sojourn on the sea bed, it computed and displayed the movement of the Sun, the Moon and possibly the planets around Earth, and predicted the dates of future eclipses. It's one of the most stunning artefacts we have from classical antiquity.

No earlier geared mechanism of any sort has ever been found. Nothing close to its technological sophistication appears again for well over a millennium, when astronomical clocks appear in medieval Europe. It stands as a strange exception, stripped of context, of ancestry, of descendants.

Considering how remarkable it is, the Antikythera Mechanism has received comparatively scant attention from archaeologists or historians of science and technology, and is largely unappreciated in the wider world. A virtual reconstruction of the device, published by Mike Edmunds and his colleagues in this week's *Nature* (see page 587), may help to change that. With the help of pioneering three-dimensional images of the fragments' innards, the authors present something close to a complete picture of how the device worked, which in turn hints at who might have been responsible for building it.

But I'm also interested in finding the answer to a more perplexing question — once the technology arose, where did it go to? The fact

that such a sophisticated technology appears seemingly out of the blue is perhaps not that surprising — records and artefacts from 2,000 years ago are, after all, scarce. More surprising, to an observer from the progress-obsessed twenty-first century, is the apparent lack of a subsequent tradition based on the same technology — of ever better clockworks spreading out round the world. How can the capacity to build a machine so magnificent have passed through history with no obvious effects?

Astronomic leaps

To get an idea of what the mechanism looked like before it had the misfortune to find itself on a sinking ship, I went to see Michael Wright, a curator at the Science Museum in London for more than 20 years and now retired. Stepping into Wright's workshop in Hammersmith is a little like stepping into the workshop where H. G. Wells' time machine was made.

Every inch of floor, wall, shelf and bench space is covered with models of old metal gadgets and devices, from ancient Arabic astrolabes to twentieth-century trombones. Over a cup of tea he shows me his model of the Antikythera Mechanism as it might have been in his pomp. The model and the scholarship it embodies have consumed much of his life (see 'Raised from the depths').

The mechanism is contained in a square wooden case a little smaller than a shoe-box. On the front are two metal dials (brass, although the original was bronze), one inside the other, showing the zodiac and the days of the year. Metal pointers show the positions of the Sun, the Moon and five planets visible to the naked eye. I turn the wooden knob on the side of the box and time passes before my eyes: the Moon makes a full revolution as the Sun inches just a twelfth of the way around the dial. Through a window near the centre of the dial peeks a ball painted half black and half white, spinning to show the Moon's changing phase.

On the back of the box are two spiral dials, one above the other. A pointer at the centre of each traces its way slowly around the spiral groove like a record stylus. The top dial, Wright

explains, shows the Metonic cycle — 235 months fitting quite precisely into 19 years. The lower spiral, according to the research by Edmunds and his colleagues, was divided into 223, reflecting the 223-month period of the Saros cycle, which is used to predict eclipses.

To show me what happens inside, Wright opens the case and starts pulling out the wheels. There are 30 known gear-wheels in the Antikythera Mechanism, the biggest taking up nearly the entire width of the box, the smallest less than a centimetre across. They all have triangular teeth, anything from 15 to 223 of them, and each would have been hand cut from a single sheet of bronze. Turning the side knob engages the big gear-wheel, which goes around once for every year, carrying the date hand. The other gears drive the Moon, Sun and planets and the pointers on the Metonic and Saros spirals.

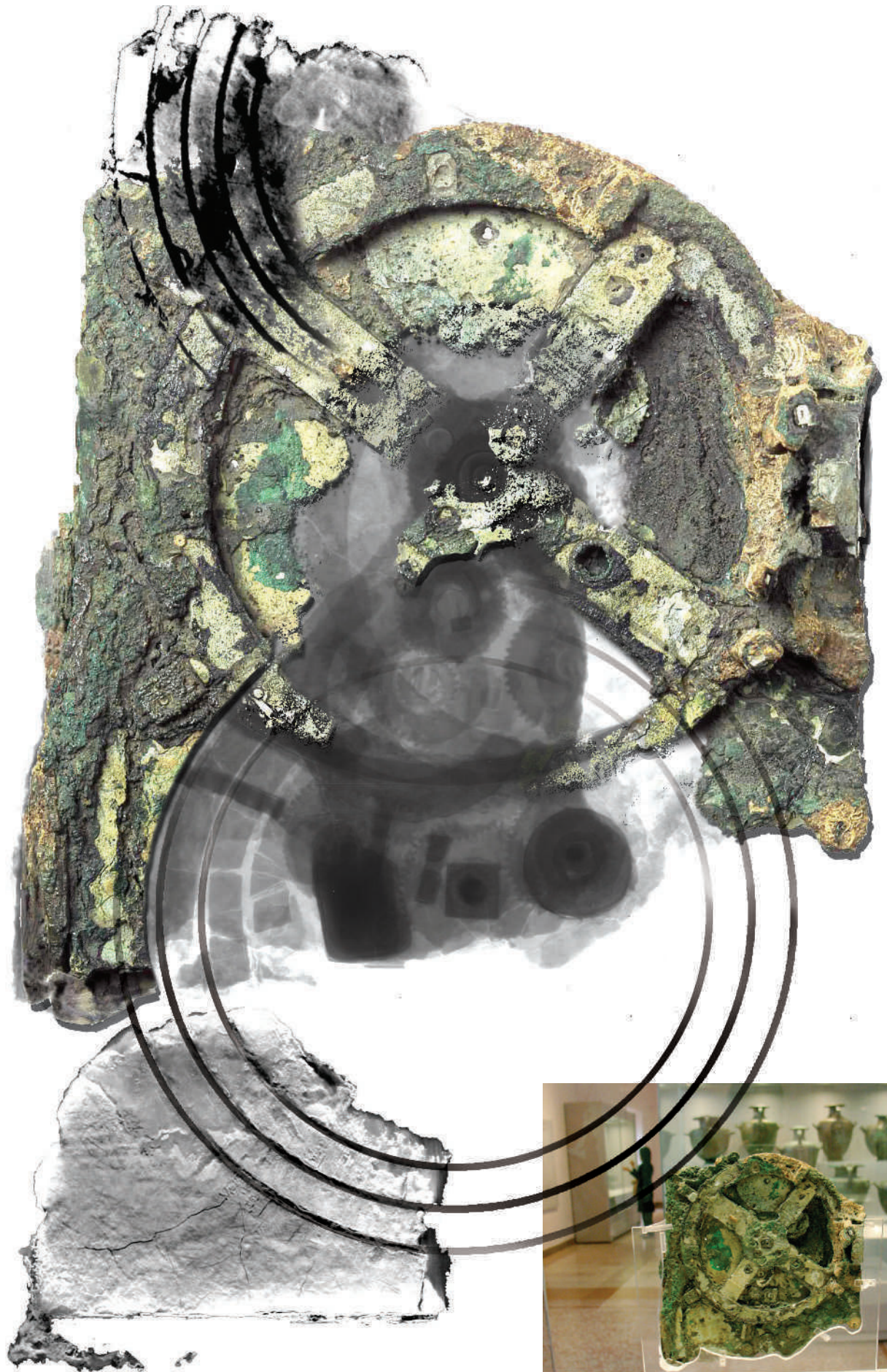
To see the model in action is to want to find out who had the ingenuity to design the original. Unfortunately, none of the copious inscriptions is a signature. But there are other clues. Coins found at the site by Jacques Cousteau in the 1970s have allowed the shipwreck to be dated sometime shortly after 85 BC. The inscriptions on the device itself suggest it might have been in

use for at least 15 or 20 years before that, according to the Edmunds paper.

The ship was carrying a rich cargo of luxury goods, including statues and silver coins from Pergamon on the coast of Asia Minor and vases in the style of Rhodes, a rich trading port at the time. It went down in the middle of a busy shipping route from the eastern to western Aegean, and it seems a fair bet that it was heading west for Rome, which had by that time become the dominant power in the Mediterranean and had a ruling class that loved Greek art, philosophy and technology.

The Rhodian vases are telling clues, because Rhodes was the place to be for astronomy in the first and second centuries BC. Hipparchus, arguably the greatest Greek astronomer, is thought to have worked on the island from around 140 BC until his death in around 120 BC. Later the philosopher Posidonius set

"It's a popular notion that technological development is a simple progression. But history is full of surprises."
— François Charette



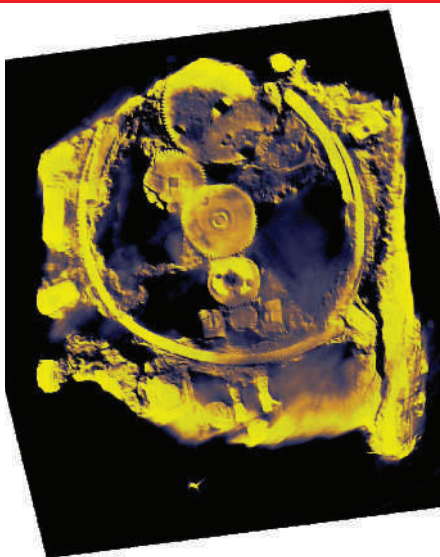
J. MARCHANT/ANTIKYTHERA MECHANISM RESEARCH PROJECT

up an astronomy school there that continued Hipparchus' tradition; it is within this tradition that Edmunds and his colleagues think the mechanism originated. Circumstantial evidence is provided by Cicero, the first-century BC Roman lawyer and consul. Cicero studied on Rhodes and wrote later that Posidonius had made an instrument "which at each revolution reproduces the same motions of the Sun, the Moon and the five planets that take place in the heavens every day and night". The discovery of the Antikythera Mechanism makes it tempting to believe the story is true.

And Edmunds now has another reason to think the device was made by Hipparchus or his followers on Rhodes. His team's three-dimensional reconstructions of the fragments have turned up a new aspect of the mechanism that is both stunningly clever and directly linked to work by Hipparchus.

One of the wheels connected to the main drive wheel moves around once every nine years. Fixed on to it is a pair of small wheels, one of which sits almost — but not exactly — on top of the other. The bottom wheel has a pin sticking up from it, which engages with a slot in the wheel above. As the bottom wheel turns, this pin pushes the top wheel round. But because the two wheels aren't centred in the same place, the pin moves back and forth within the upper slot. As a result, the movement of the upper wheel speeds up and slows down, depending on whether the pin is a little farther in towards the centre or a little farther out towards the tips of the teeth (see illustration on page 551).

The researchers realized that the ratios of the gear-wheels involved produce a motion that



Inside out: computer tomography of the main fragment allowed for accurate modelling.

closely mimics the varying motion of the Moon around Earth, as described by Hipparchus. When the Moon is close to us it seems to move faster. And the closest part of the Moon's orbit itself makes a full rotation around the Earth about every nine years. Hipparchus was the first to describe this motion mathematically, working on the idea that the Moon's orbit, although circular, was centred on a point offset from the centre of Earth that described a nine-year circle. In the Antikythera Mechanism, this theory is beautifully translated into mechanical form. "It's an unbelievably sophisticated idea," says Tony Freeth, a mathematician who worked out most of the mechanics for Edmunds' team. "I don't know how they thought of it."

"I'm very surprised to find a mechanical representation of this," adds Alexander Jones, a historian of astronomy at the University of Toronto, Canada. He says the Antikythera Mechanism has had little impact on the history of science so far. "But I think that's about to change. This was absolutely state of the art in astronomy at the time."

Wright believes that similar mechanisms modelled the motions of the five known planets, as well as of the Sun, although this part of the device has been lost. As he cranks the gears of his model to demonstrate, and the days, months and years pass, each pointer alternately lags behind and picks up speed to mimic the astronomical wanderings of the appropriate sphere.

Greek tragedy

Almost everyone who has studied the mechanism agrees it couldn't have been a one-off — it would have taken practice, perhaps over several generations, to achieve such expertise. Indeed, Cicero wrote of a similar mechanism that was said to have been built by Archimedes. That one was purportedly stolen in 212 BC by the Roman general Marcellus when Archimedes was killed in the sacking of the Sicilian city of Syracuse. The device was kept as an heirloom in Marcellus' family: as a friend of the family, Cicero may indeed have seen it.

So where are the other examples? A model of the workings of the heavens might have had value to a cultivated mind. Bronze had value for everyone. Most bronze artefacts were eventually melted down: the Athens museum has just ten major bronze statues from ancient Greece, of which nine are from shipwrecks. So in terms

Raised from the depths

In 1900 a party of Greek sponge divers sought shelter from a storm in the lee of the barren, rocky islet of Antikythera. Once the winds had eased, Elias Stadiatis dived 42 metres to a rocky shelf to look for late additions to his hard-earned haul. Instead of sponges nestled on the sea bed, the shape of a great ship loomed out of the blue. After grabbing the larger-than-life arm of a bronze figure as proof of his find, he returned to the surface to inform his companions. The Antikythera wreck was to yield a stunning collection of bronze and marble statues, pottery, glassware, jewellery and coins; it was also to claim the life of one of the divers, not yet aware of the risk of the bends when diving with an oxygen hose.

As busy museum staff struggled to piece together statues and vases, a formless, corroded lump of bronze and wood lay unnoticed. But as the wood dried and

shrivelled, the lump cracked open, and on 17 May 1902, archaeologist Valerios Stais noticed that there were gear-wheels inside.

The gears elicited interest, but it was not until investigations delved beneath the surface that the box started to yield its secrets. The British science historian Derek de Solla Price and the Greek nuclear physicist Charalambos Karakalos made X- and gamma-ray images of the fragments in 1971. Karakalos and his wife Emily painstakingly counted the visible teeth; in 1974 Price published a heroic 70-page account of the machine (*D. de S. Price Trans. Am. Phil. Soc., New Ser.* 64, 1–70; 1974).

"Price really put the mechanism on the map," says Tony Freeth, co-author of a new reconstruction of the device (see page 587). "He understood the essence of what it was — an astronomical computer." But Price massaged some of the

data (much to the annoyance of Karakalos and his wife), and his reconstruction was unnecessarily complicated — perhaps too complicated for historians and archaeologists. They largely ignored Price's work, and he died in 1983.

That same year, a Lebanese man walked into the Science Museum in London with the pieces of another ancient mechanism in his pocket.

Curator Michael Wright realized the device was a Byzantine sundial from the sixth century AD, which also contained a simple geared mechanism that drove pointers showing the position of the Moon and Sun in the sky. Studying the astronomically enhanced sundial led Wright to Price's treatment of the Antikythera Mechanism, in which he saw serious holes.



of the mechanism, “we’re lucky we have one”, points out Wright. “We only have this because it was out of reach of the scrap-metal man.”

But ideas cannot be melted down, and although there are few examples, there is some evidence that techniques for modelling the cycles in the sky with geared mechanisms persisted in the eastern Mediterranean. A sixth-century AD Byzantine sundial brought to Wright at the Science Museum has four surviving gears and would probably have used at least eight to model the positions of the Sun and Moon in the sky. The rise of Islam saw much Greek work being translated into Arabic in the eighth and ninth centuries AD, and it seems quite possible that a tradition of geared mechanisms continued in the caliphate. Around AD 1000, the Persian scholar al-Biruni described a “box of the Moon” very similar to the sixth-century device. There’s an Arabic-inscribed astrolabe dating from 1221–22 currently in the Museum of the History of Science in Oxford, UK, which used seven gears to model the motion of the Sun and Moon.

But to get anything close to the Antikythera Mechanism’s sophistication you have to wait until the fourteenth century, when mechanical clockwork appeared all over western Europe. “You start to get a rash of clocks,” says Wright. “And as soon as you get clocks, they are being used to drive astronomical displays.” Early examples included the St Albans clock made by Richard Wallingford in around 1330 and a clock built by Giovanni de’Dondi a little later in Padua, Italy, both of which were huge astronomical display pieces with elaborate gearing behind the main dial to show the position of the Sun, Moon, planets and (in the case of the

Model success:
Michael Wright
devoted his life
to decoding and
replicating the
Antikythera
Mechanism.



A. WRIGHT

Padua clock) the timing of eclipses. The time-telling function seems almost incidental.

It could be argued that the similarities between the medieval technology and that of classical Greece represent separate discoveries of the same thing — a sort of convergent clockwork evolution. Wright, though, favours the idea that they are linked by an unbroken tradition: “I find it as easy to believe that this technology survived unrecorded, as to believe that it was reinvented in so similar a form.” The timing of the shift to the West might well have been driven by the fall of Baghdad to the Mongols in the thirteenth century, after which much of the caliphate’s knowledge spread to Europe. Shortly after that, mechanical clocks appeared in the West, although nobody knows exactly where or how. It’s tempting to think that some mechanisms, or at least the ability to build

them, came west at the same time. As François Charette, a historian of science at Ludwig Maximilians University in Munich, Germany, points out, “for the translation of technology, you can’t rely solely on texts.” Most texts leave out vital technical details, so you need skills to be transmitted directly.

But if the tradition of geared mechanisms to show astronomical phenomena really survived for well over a millennium, the level of achievement within that tradition was at best static. The clockwork of medieval Europe became more sophisticated and more widely applied fairly quickly; in the classical Mediterranean, with the same technology available, nothing remotely similar happened. Why didn’t anyone do anything more useful with it in all that time? More specifically, why didn’t anyone work out earlier what the gift of hindsight seems to make

Wright ended up working with Allan Bromley, a computer scientist at Sydney University in Australia who had become interested in the Antikythera Mechanism at around the same time. Bromley wanted to study the machine with X-ray tomography, which assembles a sheaf of cross-sections of its subject. As the fragments could not be moved from the museum, and Bromley didn’t have the money to ship a tomography machine to Athens, Wright used his tool-making skills to build a crude tomograph *in situ*. The two researchers took around 700 images of the fragments, and Wright has been working on a reconstruction that supercedes Price’s ever since.

In the meantime, Mike Edmunds, an astrophysicist at Cardiff University, UK, and his friend Tony Freeth, a mathematician-turned-film-maker living in London, decided the mechanism would make a fantastic subject for a



Derek de Solla Price tried to undo the Antikythera Mechanism’s secrets.

documentary. But their efforts soon turned to discovering more about how the device worked. They contacted Hewlett-Packard, which had developed a method for reading eroded cuneiform tablets that involved building up a composite computer image from pictures

taken under light from a wide variety of directions, to reveal more of the inscriptions. They enlisted experts in computer-assisted tomography from British firm XTEC, which developed a new machine just for the Antikythera project.

In autumn 2005, the Hewlett-Packard equipment and all 12 tonnes of XTEC’s machinery were shipped to the museum. The results have allowed the team to confirm many of Wright’s ideas, and extend them. “My main fear initially was that we’d throw all this technology at it and we wouldn’t do more than dot the i’s and cross the t’s,” says Freeth. “But we got more out of it than I dared hope.”

One major new result came as much from chance as from technology; a key section of a dial found sitting unnoticed in the museum’s store room helped reveal that one of the dials was used to predict eclipses. Another big discovery was the identification of a ‘pin and slot’ mechanism to model

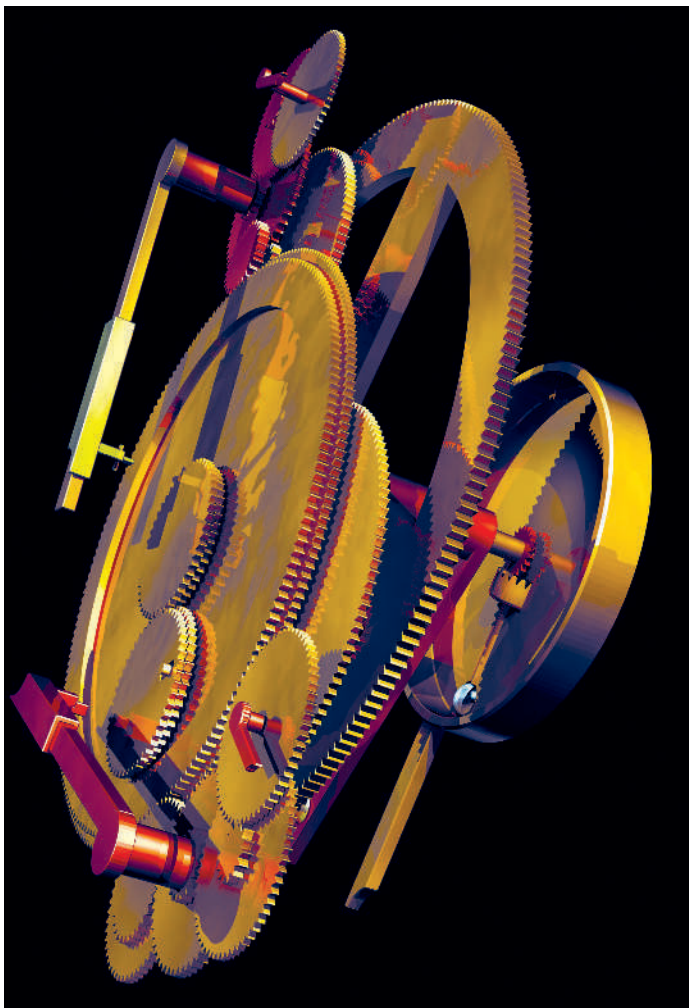
the varying speed of the Moon through the sky (see main story).

The inscriptions are also revealing novelties, although deciphering them is hard work: some of them are less than 2 millimetres high, and there are no spaces to show where each word starts and finishes. Agamemnon Tselikas, director of the Centre for History and Palaeography in Athens, spent a concentrated three months trying to decipher the wording, working from late at night into the early hours of the morning: “I needed the silence.”

So far Tselikas and his colleague Yanis Bitsakis have more than doubled the number of legible characters on the mechanism, which seem to form a manual that explains how the mechanism was to be used. It takes “imagination and intuition” to decipher the inscriptions, says Tselikas. “We are just starting to penetrate the mentality of the user of this machine.”

J.M.

M. KIRK



Gearing up: this reconstruction shows, among other things, the offset wheels of the Moon's nine-year cycle (lower left). A labelled diagram is on page 551.

obvious — that clockwork would be a good thing to make clocks with?

Serafina Cuomo, a historian of science at Imperial College, London, thinks that it all depends on what you see as 'useful'. The Greeks weren't that interested in accurate timekeeping, she says. It was enough to tell the hour of the day, which the water-driven clocks of the time could already do fairly well. But they did value knowledge, power and prestige. She points out that there are various descriptions of mechanisms driven by hot air or water — and gears. But instead of developing a steam engine, say, the devices were used to demonstrate philosophical principles. The machines offered a deeper understanding of cosmic order, says David Sedley, a classicist at the University of Cambridge, UK. "There's nothing surprising about the fact that their best technology was used for demonstrating the laws of astronomy. It was deep-rooted in their culture."

Another, not mutually exclusive, theory is that devices such as the Antikythera Mechanism were signifiers of social status. Cuomo points out that demonstrating wondrous devices brought social advancement. "They were trying to impress their peers," she says. "For them, that was worth doing." And the Greek elite was not the only potential market. Rich Romans were eager for all sorts of Greek

sophistication — they imported philosophers for centuries.

Seen in this light, the idea that the Antikythera Mechanism might be expected to lead to other sorts of mechanism seems less obvious. If it already embodied the best astronomy of the time, what more was there to do with it? And status symbols do not follow any clearly defined arc of progress. What's more, the idea that machines might do work may have been quite alien to slave-owning societies such as those of Ancient Greece and Rome. "Perhaps the realization that you could use technology for labour-saving devices took a while to dawn," says Sedley.

There is also the problem of power. Water clocks are thought to have been used on occasion to drive geared mechanisms that displayed astronomical phenomena. But dripping water only provides enough pressure to drive a small number of gears, limiting any such display to a much narrower scope than that of the Antikythera Mechanism, which is assumed to have been handcranked. To make the leap

to mechanical clocks, a geared mechanism needs to be powered by something other than a person; it was not until medieval Europe that clockwork driven by falling weights makes an appearance.

Invention's evolution

Bert Hall, a science historian at the University of Toronto in Canada, believes a final breakthrough towards a mechanical weight drive might have come about almost by accident, by adapting a bell-ringing device. A water clock could have driven a hammer or weight mechanism swinging between two bells as an alarm system, until someone realized that the weight mechanism would be a more regular way of driving the clock in the first place. When the new way to drive clocks was discovered, says Hall, "the [clockwork] technology came rushing out of the wings into the new tradition".

Researchers would now love further mechanisms to be unearthed in the historical record. "We hope that if we can bring this to people's attention, maybe someone poking around in their museum might find something, or at least a reference to something," says Edmunds. Early Arabic manuscripts, only a fraction of which have so far been studied, are promising to be fertile ground for such discoveries.

Charette also hopes the new Antikythera reconstruction will encourage scholars to take the device more seriously, and serve as a reminder of the messy nature of history. "It's still a popular notion among the public, and among scientists thinking about the history of their disciplines, that technological development is a simple progression," he says. "But history is full of surprises."

In the meantime, Edmunds' Antikythera team plans to keep working on the mechanism — there are further inscriptions to be deciphered and the possibility that more fragments could be found.

This week the researchers are hosting a conference in Athens that they hope will yield fresh leads. A few minutes' walk from the National Archaeological Museum, Edmunds' colleagues from the University of Athens, Yanis Bitsakis and Xenophon Moussas, treat me to a dinner of aubergine and fried octopus, and explain why they would one day like to devote an entire museum to the story of the fragments.

"It's the same way that we would do things today, it's like modern technology," says Bitsakis. "That's why it fascinates people." What fascinates me is that where we see the potential of that technology to measure time accurately and make machines do work, the Greeks saw a way to demonstrate the beauty of the heavens and get closer to the gods. ■

Jo Marchant is *Nature's* News Editor.

"I find it as easy to believe that this technology survived unrecorded, as to believe that this was reinvented in so similar a form." — Michael Wright



THE COST OF LEAFING

Understanding the trade-offs involved for plants making leaves promises fresh insights on every scale from the plant to the planet, finds **John Whitfield**.

There are about 250,000 different types of plant — and almost as many types of leaf, from blades of grass through downy beech leaves to the needles of cedars or the fronds of palms. But the differences hide a basic similarity, says ecologist Ian Wright of Macquarie University in Sydney, Australia. “Within any habitat, you’ll find that each square centimetre of leaf will process a roughly similar amount of carbon per unit area over its lifespan.”

Wright and his colleagues have discovered that most of the variation in the physical and biochemical properties of leaves can be represented on a single axis running, to put it crudely, from quick and juicy to slow and tough. They call the axis the ‘worldwide leaf economics spectrum’¹, and it embodies many of the trade-offs that govern how plants deploy their resources within the limits that physics places on biological possibility.

According to the Thomson in-cites website, the work is the second-most cited paper on ecology and the environment published in the past two years. It has attracted the attention of everyone, from plant physiologists studying how leaves work to biogeochemists looking at the cycling of nutrients on a global scale. In part, the paper is so popular because of the size and scope of the database that underlies the work; but the popularity also reflects the intellectual excitement that surrounds the discovery that so much

can be explained by so little. This has given some ecologists hope that by looking at the large-scale patterns in how organisms work, they can gain a general understanding of why species live where they do, and why some are common and others are rare. Such findings are not of purely academic interest: climate researchers are using them to improve their models of the consequences of global warming.

From rainforest to tundra

The spectrum was discovered by compiling and analysing a database of the leaf biology of 2,548 species from 175 locations, ranging from tropical rainforests to the Arctic tundra. For every species, six key traits were recorded: leaf mass per unit area (affected by a leaf’s thickness and tissue density); nitrogen and phosphorus content; the lifespan of the leaves; and the rates of photosynthesis and respiration. Wright and his colleagues compiled the data from their own and other people’s studies — “much of it was languishing in bottom drawers”, he says.

When the database was analysed statistically, its first principal component — the most informative way of looking at all the traits using a single axis — turned out to account for 74% of all the variation in all six traits. This principal component is the worldwide leaf economics

spectrum. It ranges from leaves that are cheap to manufacture, highly active (ie, they have a high nitrogen content and photosynthetic rate) and short-lived, to those that are expensive and less active but make up for that in durability. In terms of an English woodland, think of going from flimsy short-lived beech leaves to tough old holly.

One definition of economics is the study choice under the constraint of scarcity, and the narrow range of choices in the leaf economics spectrum provides a vivid illustration of the various scarcities that dominate plants’ lives. The fact that all leaves lie fairly close to the axis of the spectrum shows that, despite the vast diversity of foliage produced over hundreds of millions of years of evolution, plants have little room for manoeuvre in

how they build their leaves. “Most textbooks of ecology project the idea that there’s an almost infinite diversity of organisms,” says plant ecologist Philip Grime of the University of Sheffield, UK. “But if you look at the core biology of what organisms do with resources, you find severe constraints and trade-offs.”

One tack taken in following up on this ground-breaking work has been to ask what the constraining trade-offs are. When Wright and ecologist Bill Shipley of Sherbrooke University

“By looking at plant traits, you can go from physiology up to ecosystem functioning.” — Bill Shipley

J. PARDOE/ALAMY

SOME OF THE 250,000 LEAF OPTIONS AVAILABLE

L. KENNEDY/CORBIS; B. LEGLER; J. PEACH/IMAGEQUESTMARINE.COM

Cheap

Purple lupine (wild *Lupinus* sp.)Water sedge (*Carex aquatilis*)Norway maple (*Acer platanoides*)Post oak (*Quercus stellata*)

in Quebec, Canada, used a mathematical technique that can find the direction of causality in the correlations², they found that “there was no way to explain the correlations based only on the measurements in the data set”, says Shipley. Some not yet measured factor was at work.

The missing link

Shipley thinks that the missing fundamental trade-off is between the volume of a leaf's cell and the thickness of its cell wall. Leaves at the tough end of the spectrum have cells like — well, like cells: small spaces surrounded by thick, solid walls. Cell walls require a lot of materials to make and do not photosynthesize. Sturdier cell walls make the leaves more opaque and less permeable to gases, hindering the capture of light and carbon dioxide. But these leaves also live longer, and the investment in the cell wall is thus paid off with a low rate of photosynthesis that can be maintained for a very long time.

Another way to take the work forward is to include some aspects of a leaf's size and shape. The factors measured in the spectrum are usually expressed per unit mass. But Charles Price of the University of Arizona, Tucson thinks that a leaf's absolute size is also important. He has sought to explain how a leaf's surface area and mass are linked in terms of the geometry of the vein network that carries resources to, and away from, its cells³. Large networks supply cells more slowly: this, he predicts, will cause large leaves to photosynthesize more slowly, per unit area, than small ones, and might be another reason why leaves with a large mass per area photosynthesize more slowly.

Perhaps the most striking aspect of the trade-offs to be investigated further, though, is that different plants in the same places make very different ‘decisions.’ “The dogma was that plants differed between places that are dry and wet, and those that have low and high levels of nutrients,” says Wright. “But frequently, the range of trait values within any one habitat is as large or larger than that seen between sites.” This illustrates, he

says, that the various ways of making a leaf are equally good routes to the same place. Variation within sites will also result from small-scale changes in soil, shade and moisture, and from the fact that a plant's best strategy depends on what everything else is doing — in a wood full of fast-growers, tolerating shade might be better than trying to compete directly.

Climate, though, does play a part in the shape of the spectrum. In the coldest places, doubling the leaf mass per area lengthens the lifespan by five times, whereas for the hottest sites, the same doubling lengthens lifespan by only 2.5 times.



Vintage years: the same vineyard after 44 (top) and 42 (middle) years of abandonment, and another (bottom) after just five years.

About 20% of the variance not captured in the basic leaf economics index is explained once climate is added to the equation⁴.

Onwards and upwards

The implications of a leaf's-eye view go beyond the plants themselves. Leaves are the gateways through which energy enters ecosystems, and choices made on the leaf economics spectrum affect detritivores, herbivores and, indirectly, carnivores. Plants with short-lived, rapidly photosynthesizing leaves are less tolerant of self-shading, and so have broader, flatter canopies than those with tougher, slow-growing leaves, such as conifers, which tend to grow more compactly. And the litter from plants that are fast-living decomposes more quickly than the tough foliage of the slow-coaches, controlling the speed at which nutrients cycle through the ecosystem. This sort of thinking is bolstering some ecologists in their belief that the study of traits is central to many aspects of their discipline⁵.

To try to understand things such as what determines the number of species that can coexist in a place, how numerous each species is, and how productive the system as a whole is, ecologists have traditionally looked at what species are present, how they interact, and how their abundance affects that of the others. Such an approach is an extension of ecology's roots in natural history, says Brian McGill of McGill University in Montreal, Canada. “People become ecologists because they love to go outdoors and look at the woods. They get attached to putting names on things, and get focused on knowing lots about particular organisms.”

But this approach soon becomes intractably knotty, as the number of possible interactions between species rises geometrically with the number of species. “We don't have the capacity to learn as much as we need to know by studying one species at a time. Studying interactions between species, and then trying to build that up, hasn't panned out. It's too complicated,” explains McGill.

Red bloodwood (*Corymbia gummiifera*)Woollybush (*Adenanthos cygnorum*)Scots pine (*Pinus sylvestris*)Monkey-puzzle tree
(*Araucaria araucana*)

Expensive

Seeing ecological communities in terms of their constituent traits, rather than their species, might be less intuitive. But it makes measuring and comparing places a lot easier, says McGill. The trick is to find traits that are closely related to evolutionary fitness. The leaf economics spectrum is a prime example of this. Another example is the spectrum between plants that make a lot of small seeds and those that make a few big ones. Wright is now looking for other important axes, such as between plants that use water profligately to grow quickly, and those that are slow-growing but intolerant to drought.

Once researchers think they have identified an axis of variation, they can see, for example, whether plants cluster tightly at one point on the axis, suggesting that they have evolved similar responses to a common environmental challenge, or whether they spread out along its length as they do on the leaf economics spectrum, suggesting that they have adopted strategies to avoid competing with one another.

In October, Shipley and his colleagues published a study that tested how predictive such an approach might be⁶. Working in 12 French vineyards abandoned 2–42 years ago, they looked at the ways the traits of the plants change as the vineyards revert to an uncultivated state. As perennials replaced weedy species, the height and mass of the plants increased, and the number of seeds per plant dropped. The leaves of the perennials also had a greater mass per area, in a shift towards the slow/tough end of the economic spectrum that seemed to go along with the community's ageing.

From a knowledge of the average trait values for all 30 species looked at in the study and the age of each field, the researchers then predicted the abundance of each species in each vineyard by borrowing a technique from statistical mechanics and assuming that resources were spread between the species as randomly as possible. They predicted their number of species in each plot and their abundance with 94% accuracy. "By looking at plant traits, you can

go from physiology all the way up to ecosystem functioning," says Shipley. "The potential that I see is in finally being able to integrate different levels of plant ecology together."

A species'-eye view might still reveal things about ecosystems that looking at traits cannot, says David Tilman of the University of Minnesota in St Paul. Plots with more species, for example, are more productive, and recover more quickly from perturbation than less diverse places⁷. Part of this trend is that more species mean a greater diversity of traits, Tilman says, but interactions between species are also important. "Looking at traits won't let us ignore the number of species. Both are equally important determinants of how an ecosystem functions."

Predicting the future

But the role a trait-based approach could play in prediction is exciting some of the scientists studying the greatest environmental problem of the age at the largest possible scale—the planet-wide effects of global warming. A wide array of feedbacks ties the terrestrial biosphere into the changes in the atmosphere's carbon dioxide content. Making sense of these relationships requires models that predict what plants to expect where, on the basis of environmental factors such as temperature and water availability.

Today's models typically classify vegetation as belonging to one of 5–20 groups — evergreen, deciduous, woody, herbaceous and so on. Trait values, such as leaf longevity and photosynthetic rate, are then assigned to this group as a whole. This summer, researchers working on plant traits and those working on the interaction between plants and climate met in Sydney to try to see how a 'trait-first' approach to modelling might improve things.

"Either we watch the world change, or we gain a truly predictive knowledge of nature." — David Tilman

It turns out that models that replace a fixed trait value assigned to each vegetation type with a spread of values taken from the leaf economics spectrum give values of plant productivity closer to those seen on the ground. They should thus yield more accurate predictions of future change,

particularly at a regional or continental scale, says modeller Ian Woodward of the University of Sheffield. "Looking at Amazonia, we tend to find that incorporating these data into our models leads to changes in leaf longevity and primary productivity of about 10–15%. That's a significant difference, and an exciting eye-opener," he says.

In general, says Tilman, if we are to understand and mitigate the effects of processes such as climate change, nitrogen deposition, species introductions and habitat loss, ecology needs to home in on the most important ways in which species differ. "These broad global patterns greatly simplify what we need to understand about species. They let us sort through a hopeless morass of information, and go from conceptual models to real models that say how ecosystems will respond to environmental change," he says.

"If we can't be mechanistic and predictive, we'll be unable to provide society with explanations of the impacts of our actions. Either we watch the world change, and explain it in retrospect, or we gain a truly predictive knowledge of nature."

John Whitfield is a freelance writer based in London, UK.

1. Wright, I. J. *et al.* *Nature* **428**, 821–827 (2004).
2. Shipley, B., Lechowicz, M. J., Wright, I. & Reich, P. B. *Ecology* **87**, 535–541 (2006).
3. Price, C. A. & Enquist, B. J. *Ecology* (in the press).
4. Wright, I. J. *et al.* *Glob. Ecol. Biogeogr.* **14**, 411–421 (2005).
5. McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. *Trends Ecol. Evol.* **21**, 178–185 (2006).
6. Shipley, B., Vile, D. & Garnier, E. *Science* **314**, 812–814 (2006).
7. Tilman, D., Reich, P. B. & Knops, J. M. *Nature* **441**, 629–632 (2006).

THE EARTH-EATERS

Research suggests that consuming soil may have more health implications than one might expect. **Trevor Stokes** sieves through the reasons why people include dirt in their diet.

At Calabash, a West African specialty grocery store in Newark, New Jersey, plastic containers line the impulse-purchase aisle next to the cash register. Hungry shoppers can choose from a wide range of traditional edible offerings, including sacks of caffeine-rich kola nuts, packets of vanilla sugar — and thumb-sized rolls of chalky clay.

In such markets, edible clay isn't a strange thing to find. Joyce Corleley, a Ghanaian customer in the store, says she ate fire-cured clay, known as *shra*, during two of her pregnancies. The dirt became a food substitute that gave her the illusion of eating, without the fear of vomiting food during morning sickness.

Corleley says she doesn't eat clay for medical reasons. "It doesn't give you energy, it doesn't make the baby grow faster, it doesn't make the baby healthy," she says. "It doesn't have any benefit, it's just clay." Several recent scientific studies bear her out — but others suggest otherwise.

Anthropologists and biologists have long tried to explain geophagia, a practice named after the Greek words for earth-eater. Some researchers say that eating clay helps supplement a person's diet with much-needed minerals. Others argue that the stomach's acidity would remove any of the nutritional benefits. Still others have been searching for any potential evolutionary advantage to the practice, which falls under the general term of *pica* — the consumption of non-nutritive substances ranging from paper to cigarette butts.

Dirty past

Despite these recent studies, geophagia has received relatively little attention from the research community. "It's one of those topics that, paradoxically, everyone sees as being important, but nobody actually has it as their specialized area of research," says Jeya Henry, a human nutritionist at Oxford Brookes University, UK, and one of the editors of the forthcoming overview book *Consuming the Inedible*¹.

Records of geophagia exist in all corners of the world, stretching back to 1800 BC in Sumeria, Egypt and China, according to Rudolph Reinbacher, a historian in California². For instance, at least 2,000 years ago, he says, Greek markets sold *terra sigillata*, clay specially minted into "health" coins that had supposed medicinal properties. "People must have felt there was something to it," he says, "so they ate it."

What that 'something' is remains in dispute.



Chewing the mud: A Pemban woman collects clay for the practice of geophagia.

According to one leading theory, the dirt serves as a sort of multivitamin. Clay, after all, consists of a lattice of silicon dioxide and aluminium oxide — similar to pill filler — mixed with trace minerals such as calcium, iron and zinc.

In the 1990s, nutrition researcher Susanne Aufreiter of the University of Toronto, Canada, decided to test this 'nutrition hypothesis'. Her team analysed clays that people had eaten during the latter part of the twentieth century in various parts of the world: in China, eaten during famines; in North Carolina, eaten for general health; and in Zimbabwe, eaten to treat diarrhoea. Each of the clays contains levels of trace minerals that could supplement a poor diet, Aufreiter says³.

Earth eating might also provide other ben-

efits, she adds. Vegetables wouldn't have been scrubbed the way they are today, and Aufreiter argues that residual earth would have added roughage to the diet. "We didn't evolve with this wonderful hygienic life that we have."

But chemical analyses on clay can miss the nuances of the digestive system through which clay moves. Digested clay may not only hold on to trace elements, but may actually leach them from intestinal tracts, other work suggests.

At Kingston University in Surrey, UK, a team led by soil scientist Peter Hooda created a slurry of clay and simulated gastric acids and nutrient solution, and incubated it at body temperature to simulate conditions in the stomach. The team found that nutrients can become tightly bound to the lattice structure of clay particles, resulting



Food supplements: a variety of samples of earth that are consumed in Zanzibar.

in lower levels of iron, zinc and copper in the slurry⁴. And the higher the acidity, the more tightly the metals became bound to the clay — bad news for the nutritional hypothesis.

Eating soils rich in iron and zinc, it seems, may actually worsen malnutrition, other research suggests. At Cornell University in Ithaca, New York, nutritionist Sera Young has been trying to unravel whether geophagia or malnutrition comes first. Working on the African island of Pemba, off mainland Tanzania near Zanzibar, Young has been studying pregnant women's perceptions of iron-deficiency anaemia.

During her field-work in 2001, Young asked women in Swahili what caused anaemia. One Pamban woman responded that eating dirt from the walls of her home caused anaemia. Young thought she had misunderstood the answer, as Swahili was new to her; but no, the woman pointed and confirmed that her dirt walls caused anaemia. Other Pambans claimed the opposite — that eating earth during pregnancy actually treated their anaemia, rather than caused it.

Intrigued, Young began searching for other cases. Certain that a search of PubMed, the medical journal archive, would clear up the underlying cause of geophagia, she says: "It just was not that simple." One of the biggest difficulties is that geophagia embraces several facets of science. "You need to dabble in soil science, you can't shy away from ethnography, you have to have some epidemiological skills and some biochemical knowledge," she says. "And that's a tall order."

Young and her PhD adviser, nutritional scientist Rebecca Stoltzfus, hope to meet the challenge. In an unpublished epidemiological study of 2,500 Pambans, women who told the researchers that they ate clay had significantly lower levels of haemoglobin, and thus were anaemic. It's not clear yet, though, whether geophagia caused the anaemia, anaemia caused the geophagia, or if a third factor is causing both of them. To test this last option, Young plans to look at the elemental and mineralogical profiles of the clays, to see if there are other bioavailable minerals that could be affecting the women.

Some hints may come from work performed as far back as the late 1950s, which suggested that zinc deficiency may cause geophagia. Ananda Prasad, now a professor at Wayne State University in Detroit, Michigan, encountered several young Iranian men who suffered from stunted growth and slower sexual development. One 20-year-old man physically resembled an eight-year-old boy. All the patients ate clay every day. But when given zinc, they matured sexually and lost their desire to eat clay⁵.

Decades of subsequent work by Prasad and his colleagues showed that a lack of zinc leads to hypogeusia, a taste-diminishing disorder. Prasad reasons that hypogeusia makes dirt if not enjoyable, at least not unattractive, which reinforces the geophagia, leaches more zinc and starts the cycle all over again. And he also doesn't agree with the nutrition hypothesis: "Soils contain nutrients, that's true, that's how we get our nutrients from plants," he says. "But plants absorbing nutrients from soil is different from humans absorbing nutrients from soil."

"Plants absorbing nutrients from soil is different from humans absorbing nutrients from soil." — Ananda Prasad

Ground rules

In perhaps the only randomized, controlled study to see whether anaemia causes geophagia or vice versa, Mbiko Nchito from the University of Zambia and her colleagues have found that supplementing the diet of geophagic, anaemic children with iron did not stop them eating earth⁶. The geophagic children were also more often infected with hookworm, which causes anaemia; but that alone could not explain all the geophagic cases, as several non-infected anaemic children also ate clay.

Similarly, Young's work among Pamban women shows that the clays they eat contain no detectable levels of hookworm. So the simple explanation that hookworm-riddled clay causes nutrient deficiencies doesn't seem to pan out.

The question remains: what possible evolutionary benefits could arise from eating clay? To

address this question, Young has joined forces with Paul Sherman, a social behaviourist at Cornell University in Ithaca, New York, and Julius Lucks, a biophysics graduate student at Harvard University in Cambridge, Massachusetts. They scour popular and scientific literature for first-hand accounts of pica, in order to build up a comprehensive database of the circumstances under which it occurs. To date, they have 1,738 such accounts.

Sherman had previously studied morning sickness as a potential evolutionary adaptation, and looks at geophagia the same way. Clay could serve to help detoxify the body during pregnancy, he argues, by binding to plant toxins such as solanin in raw potatoes or nicotine in tobacco. Morning sickness is most prevalent during the first trimester of pregnancy, which is also when geophagia most commonly shows up. "So putting two and two together," he says, "perhaps there is some advantage in terms of detoxifying plants with strong secondary compounds to protect the fetus against carcinogens and mutagens." Losing nutrients to clay, Sherman argues, is probably negligible in terms of nutrition compared to the cost of throwing up the same toxins.

So far, the group's preliminary results support the idea that those most susceptible to toxins — young children and women in the early stages of pregnancy — most commonly practise geophagia. And, consistent with clay's possible detoxification role, researchers have observed that both men and women eat dirt to treat 'gastrointestinal stress', presumably cases of food poisoning. Over-the-counter stomach remedies, such as Kaopectate and Maalox, contain clay-like ingredients that settle the stomach. According to Sherman's findings, dirt-eaters self-medicate with every chunk of clay they chew.

Back at the Calabash grocery store, Joyce Corlethey picks a tiny corner of clay off a roll and gnaws it matter-of-factly. "You just put it in your mouth so that you have something in there," she explains. But once she was no longer pregnant, her experiment with clay ended. ■

Trevor Stokes is a freelance writer in New York City.

1. MacClancy, J., Macbeth, H. & Henry, J. (eds) *Consuming the Inedible: Cross-Disciplinary Approaches* (Berghahn, Oxford, 2007).
2. Reinbacher, W. R. *Healing Earths: The Third Leg of Medicine. A History of Minerals in Medicine* (1st Books Library, Bloomington, 2003).
3. Aufreiter, S. et al. *Int. J. Food Sci. Nutr.* **48**, 293–305 (1997).
4. Hooda, P. S. et al. *Sci. Total Environ.* **333**, 75–87 (2004).
5. Prasad, A. S., *Am. J. Clin. Nutr.* **53**, 403–412 (1991).
6. Nchito, M. et al. *Trans. R. Soc. Trop. Med. Hyg.* **98**, 218–227 (2004).

Islam: governments need to reform education and build a scientific culture

SIR — We read with interest your recent “Islam and science” issue focused on science in the Muslim and Arab world (*Nature* **444**, 1, 19–29 and 33–36; 2006). It addresses many of the barriers to the advancement of science in the Muslim world.

As scientists from the region — we both come from Yemen — we would like to stress that the biggest impediment to the progress of science in the Arab/Muslim world is not Islam but the absence of a scientific research culture and of national research strategies, combined with the failure of the educational systems, particularly higher education. The greatest contributions to science occurred during an era when Islam dominated the civilized world.

Research and education are inseparable: you cannot have one without the other. Higher-education institutions and centres of excellence are the driving engines of the economic growth and competitiveness of the knowledge-based economies in the West. The number of research-based PhD programmes is limited or non-existent in many Muslim/Arab countries.

At the core of the problem in the Muslim/Arab world is the existence of, primarily, two forms of educational institution: government and for-profit private institutions. Neither system is compatible with the development of a world-standard scientific infrastructure.

The government institutions have funding, but no financial independence or control of their resources. Their faculty members and administrators are appointed on political grounds, not on merit. So whether the country is rich like Saudi Arabia or poor like Yemen, the result is the same: substandard education and research.

On the other hand, the number of private universities in the Arab world is rising. However, almost all are profit-driven and generally focus on small programmes that require little infrastructure.

This situation contrasts significantly to that in the West, where universities are non-profit, self-governing institutions, and governments provide partial funding but do not dictate how the money is spent.

In your Editorial “Science and the Islamists” (*Nature* **444**, 1; 2006) you correctly indicate the need for rulers to foster and seek inspiration. The terrible conflicts continuing in most Muslim countries are affecting millions of people’s basic everyday needs. Clearly, this has deterred the advancement of science in those countries and left contributions in science to other nations.

However, let us distinguish clearly

between Islam’s position on science and current practices.

The first instruction revealed to the prophet Muhammad was to read, think and ponder about God’s creation. The Koran and the record of the prophet’s practices (the sunnah and hadiths) are full of scientific concepts that inspired Muslims in the past to contribute to humanity and civilization.

The world has lost a lot because of the current state of affairs in Muslim countries. One would hope that, if the past predicts the future, then under good Islamic rule science will prevail again. Muslim scientists will then once again make pioneering contributions to science and to many other aspects of life.

In order to advance science and build knowledge-based economies, Arab and Muslim countries must:

- create a scientific culture that recognizes the contribution of scientists and values higher education and research;
- build state-of-the-art research facilities as centres of excellence;
- set up merit-based structures that foster academic freedom and promote independent scholarship;
- adopt policies that facilitate the flow of information and exchange of research materials between scientists and their counterparts around the world;
- take active approaches to reverse the brain drain of Muslim and Arab scientists and attract talent to the region, irrespective of religion, nationality or ideology;
- separate politics from scientific education.

Advancing science and building knowledge-based economies in the region will require concerted and systematic efforts by leading Muslim academics, intellectuals and policy-makers. Lessons could be learnt from the recent success stories of Asian countries such as China, South Korea and Singapore.

As directors of our respective laboratories, we are among the many Arab/Muslim scientists who have left the region and are now well-established abroad. Few are likely to abandon their position and return home at present, because the conditions that attracted them abroad do not exist in the region. In this respect, not only has the Arab/Muslim world failed to produce and retain scientists, it has also failed to capitalize on the large resources of successful and accomplished expatriate Muslim scientists.

Until the situation changes, Arab/Muslim countries should reach out and develop efficient mechanisms for mobilizing expatriate scientists and utilizing their knowledge, expertise and resources for

their home countries’ development.

Hilal A. Lashuel*, Nasser Zawia†

*Laboratory of Molecular Neurobiology and Neuroproteomics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

†Neurotoxicology and Epigenomics Laboratory, Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, Rhode Island 02881, USA

Islam: science is held back by paternalistic traditions

SIR — The News Feature “An Islamist revolution” (*Nature* **444**, 22–25; 2006) and other articles in *Nature*’s “Islam and science” special issue make interesting reading.

As a former executive director of the Pakistan Medical Research Council, I note that the current lack of capacity for science in Islamic countries is well highlighted, and the great uncertainty in predicting its future (despite some recent positive developments) is rightly underscored.

Islamic countries do not currently have the right environment or culture for science. They lack all the critical ingredients required: policies, institutions, critical mass of productive and vocal scientists and an aware and appreciative population. Despite some lip-service being paid, science and research are not valued. Therefore, it is debatable whether the form of government — secular or Islamist — is going to make much difference to the current situation in these countries.

Your News Feature rightly points out that Islam is interpreted differently in different countries and even in different parts of the same country. Therefore, religion cannot be the reason for the current science deficit — and a look at the history of Islamic-era science will support this view.

Paternalistic cultures in the Islamic countries can more reasonably be blamed. Under these cultures, inquiry and freedom of expression are actively discouraged in the home, at school, at work and in response to government policies. The capacity for critical analysis is a fundamental requirement for science, but where it has no chance of developing under lifelong suppression, how can science and research be expected to flourish? Money alone cannot address the problem: a more fundamental change in thinking and behaviour is needed.

Tasleem Akhtar

Fatima Memorial System, Shadman, Lahore 5400, Pakistan

Readers are encouraged to add their comments to the Islam and Science commentary on *Nature*’s News Blog at: http://blogs.nature.com/news/blog/2006/11/islam_and_science.html

BOOKS & ARTS

The Hungarian martians

Five of Budapest's finest changed the world in the twentieth century.

The Martians of Science

by István Hargittai

Oxford University Press: 2006. 376 pp.

£19.99, \$34.50

Arthur I. Miller

The Martians of Science tells the gripping story of five brilliant and colourful Hungarian scientists — Theodore von Kármán, John von Neumann, Leó Szilárd, Edward Teller and Eugene Wigner — who had an extraordinary impact on their profession and on world events in the twentieth century. Fritz Houtermans, a physicist who knew them all in the 1930s, once quipped that they “were really visitors from Mars”, hence the book's title. István Hargittai, a fellow Hungarian, tells their collective story for the first time, basing his research on interviews and documentation.

All five were from Budapest. They all attended its excellent high school, went to its technical university, completed their studies in Germany and eventually emigrated to America. Von Kármán, the eldest, was born in 1881, Szilárd in 1898, Wigner in 1902, von Neumann in 1903 and Teller in 1908. All were raised in middle-class, highly intellectual Jewish households. In the years after the First World War there was a dramatic increase in anti-Semitism in Hungary. With the exception of Teller's, their families converted to Christianity, which offered them some protection until the rise of Nazism in the 1930s, when converted Jews were rudely reminded of their origins. By the early 1930s, all five men were well known in America.

Von Kármán was a world-renowned expert in aerodynamics. He worked for the German air force in the First World War and continued to work for them afterwards as a consultant, even though this contravened the Treaty of Versailles. In 1930 he became director of the Guggenheim Aeronautical Laboratory at the California Institute of Technology. Air Marshall Hermann Göring invited him back to Germany in 1934 to work as a consultant, telling German critics: “I decide who is a Jew.” In the Second World War and afterwards, von Kármán made important contributions to the development



High fliers: aerodynamics expert Theodore von Kármán (above, writing) and Presidential Medal of Freedom winner John von Neumann (left, seated).



of jet aircraft and ballistic missiles.

Von Neumann studied two different subjects at two universities, completing a degree in chemical engineering at the ETH in Zurich in 1925, and a PhD in mathematics at the University of Budapest for work on axiomatic set theory. “Endearing, a genius, but not a modest man,” as a colleague recalled, von Neumann ranged over the scientific landscape. With his bottomless supply of scatological jokes, he was the life of the party. Among his lasting achievements were the design of digital computers and the development of nuclear weapons, another of his passions. He effectively began the field of game theory and contributed to economics, mathematics and quantum physics.

When Szilárd was studying at the University of Berlin, the front row of the famous physics colloquia was reserved for the likes of Albert

Einstein, Max von Laue and Max Planck. Students were supposed to sit towards the rear, but Szilárd thought otherwise. He wanted to tell Einstein about some of his recent ideas. Einstein was impressed. Together they applied for patents on a silent refrigerator run by an electromagnetic pump operating a liquid metal coolant, thus avoiding the possibility of noxious fumes.

The day after Szilárd fled Germany in March 1933, the borders closed. He took this as a lesson in life: “If you want to succeed in this world you don't have to be much cleverer than other people; you just have to be one day earlier.” He became a peripatetic scientist and for the rest of his life always kept a packed suitcase ready. Among his discoveries were the principles of a nuclear chain reaction and its potential uses in both peace and war. In 1939 he persuaded Einstein to inform President Franklin D. Roosevelt about it. Einstein's letter — virtually dictated by Szilárd — was one factor leading to the Manhattan Project and the atomic bomb. Szilárd spent much of his life raising ethical issues about the postwar role of nuclear weapons, and Hargittai believes that for this he should have shared the 1962 Nobel peace prize with Linus Pauling.

Like Szilárd, Wigner did his PhD at the

NASA

DWIGHT D. EISENHOWER LIBRARY, ABILENE

University of Berlin. He pioneered the use of group theory to explore the symmetry properties of atoms. For this he was awarded the Nobel Prize in Physics in 1963, the only one won by the martians. Strangely, Hargittai has the least to say on him — perhaps he was the least colourful.

Teller was one of Werner Heisenberg's students at the University of Leipzig and went on to make important contributions to molecular physics, astrophysics and nuclear physics. His period at Los Alamos was central to his political development. There he read Arthur Koestler's *Darkness at Noon*, which reminded him of the appalling treatment of his Russian friend and colleague Lev Landau at the hands of the NKVD, the Soviet Union's secret police, and he became a virulent anti-communist. Teller became a leading figure in the Manhattan Project and was obsessed with developing a nuclear weapon that would relegate the bombs dropped on Japan to mere firecrackers. The

result was a serious falling out with others in Los Alamos, notably Robert Oppenheimer, who had no interest in Teller's project to make a thermonuclear bomb and afterwards disagreed with its development.

Apart from Szilárd, the martians were fanatical about the cold war. Hargittai discusses this in detail while criticizing Teller's "reckless" testimony against Oppenheimer in 1953.

This is an important story that needs to be told, and Hargittai tells it well, although I would liked to have learnt more about the martians' creativity. Their politics aside, they were brilliant thinkers who were able to spot connections among apparently unconnected disciplines and thus identify fundamental problems — and then solve them. ■

Arthur I. Miller is emeritus professor of history and philosophy of science at University College London. His latest book, *Empire of the Stars: Friendship, Obsession and Betrayal in the Quest for Black holes*, was shortlisted for the Aventis Prize.

Noether and Dorothy Hodgkin (hardly a physicist), make yet another appearance, even though the sole factor that unites them — the gender that made it so hard for them to succeed — is deliberately scarcely mentioned.

The contributors were asked to submit short accounts divided into two sections: 'important contributions' and 'biography'. Many of the brisk summaries of scientific discoveries seem oddly redundant — anyone who can follow the boxed discussions of Ricci tensors, 4S or string theory does not need to read them. The editors' prescriptive format has resulted in a book packed with facts, occasionally relieved by a brief anecdote — ideal for diligent students preparing accurate but unreflective assignments.

Nevertheless, the index provides clues to more interesting stories about this book's subjects. The long list of 'firsts' reveals that it was not until 1962 that the French Academy of Science admitted a woman, Marguerite Perey, and that the first woman to receive an honorary doctorate from Princeton University was the physicist Chien-Shiung Wu in 1958. The entries under 'Nobel prize' are dominated by women said to have been unjustly passed over, including astrophysicist Jocelyn Bell Burnell, nuclear physicist Lise Meitner and chemist Agnes Pockels (strangely, Rosalind Franklin is absent, despite the presence of another X-ray crystallographer, Kathleen Lonsdale, one of the first two women elected to the Royal Society in 1945). The heading 'Nazis' — referring to the plight of Myriam Sarachik, Marietta Blau, Hertha Sponer and others — demonstrates that, contradicting the editors' desires, biographical accounts often demand discussions of discrimination.

As Byers' introduction points out, most of the contributors are practising scientists who are unused to writing history. Although they each provide a short bibliography, they have mostly omitted the many excellent books and articles written by professional historians of science. This collection would have been of more value for aspiring young women if it had provided a more nuanced appreciation of how individual scientists have been converted into exaggerated stereotypes. Curie, for instance, is often depicted as the laboratory equivalent of a domestic drudge, a selfless heroine who neglected her health and her appearance while she systematically processed tonnes of dirty pitchblende to isolate radium. Such presentations reinforce the view that female scientists are a substandard breed, neither normal women nor stellar intellects. In his foreword, Freeman Dyson perceptively criticizes the editors for not including younger physicists, who would have provided more relevant role models. He also pinpoints what remains, unfortunately, excellent advice for any ambitious woman: marry the right man. ■

Patricia Fara is at Clare College, University of Cambridge, Cambridge CB2 1TL, UK. Her recent books include *Pandora's Breeches: Women, Science and Power in the Enlightenment* (2004).

Women or just good scientists?

Out of the Shadows: Contributions of Twentieth-Century Women to Physics
edited by Nina Byers & Gary Williams

Cambridge University Press: 2006.
498 pp. £30, \$35

Patricia Fara

"I do not agree with sex being brought into science at all. The idea of 'woman and science' is completely irrelevant. Either a woman is a good scientist, or she is not." Not the provocative statement of a modern feminist, but a plea for equality voiced a century ago by Hertha Ayrton, the electrical experimenter who, in 1904, became the first woman allowed to present her own paper at the Royal Society in

London. Ayrton would presumably be furious to find herself the opening entry in *Out of the Shadows*, a chronologically arranged set of essays on female physicists.

Like Ayrton, most eminent women prefer to be remembered for their achievements, rather than their X chromosomes, but US physicists Nina Byers and Gary Williams have made womanhood an essential criterion for inclusion in this edited collection. Perhaps to fend off accusations of transgressing political correctness, they whittled down their long list of potential entries by choosing the 40 candidates who had made the most notable contributions to scientific progress. As a result, several over-familiar icons, such as Marie Curie, Emmy

Chien-Shiung Wu was the first woman to receive an honorary doctorate from Princeton University.



Understanding cancer

The Biology of Cancer

by Robert A. Weinberg

Garland Science: 2006. 796 pp.

£89.99, \$140 (hbk); £41.99, \$99 (pbk)

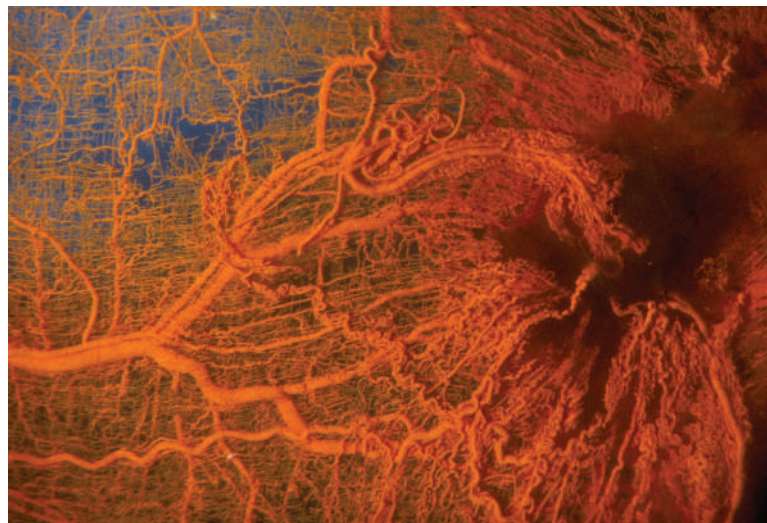
Chris Boshoff

Every cancer researcher knows 'The hallmarks of cancer', an iconic review by Douglas Hanahan and Robert Weinberg (*Cell* **100**, 57–70; 2000) that outlined the principles of tumour development: self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, induction of angiogenesis, invasion and metastasis. In *The Biology of Cancer*, Weinberg has expanded these hallmarks into an alluring text. He takes us on a journey from the pioneering days of Gregor Mendel and Peyton Rous, to the latest insights in understanding how tumours become resistant to targeted therapy.

The book comes laden with expectations, mainly a result of Weinberg's reputation as a cancer biologist and a teacher. The author has borne the weight. Each chapter unfolds as a fascinating tale, from the historical perspective, through key experiments, to interpretation and future challenges. A striking example is the section on the inhibition of pancreatic cancer and medulloblastoma growth (through targeting of the signalling protein Hedgehog), which starts with the story of sheep flocks grazing on corn lily. Their offspring developed malformations including cyclopia, which led to the isolation of cyclopamine and the eventual discovery of mutations in the *patched* and *sonic hedgehog* genes, and new analogues of cyclopamine that inhibit these. The tumour's microenvironment (including the interactions between the seed and the soil) is discussed in detail and becomes an independent hallmark of cancer, along with immunoediting, the crosstalk between tumour niche and host immunity. Weinberg's own laboratory linked the master regulators of certain processes of development (Slug and Twist) to cancer metastases. The insight obtained from processes involved in mammalian development is a recurring theme throughout the book.

In a book of this scope, some readers will inevitably feel disappointed. For example, those studying the bioenergetics of tumour cells will not find a reference to Otto Warburg or discussions about alterations in glycolytic pathways and mitochondrial respiration during tumorigenesis. Epigenetic changes in cancer are covered cursorily, and the book only touches on certain emerging fields such as the role of microRNAs in tumour development, diagnosis and classification, and the mechanisms by which autophagy (a vacuolar process of cytoplasmic degradation) contributes to oncogenesis. These are merely observations, not an indictment.

To get a good idea of the magnitude of this



Taking over: a transplanted sarcoma (on the right) has developed an unruly vascular system.

R.D. ACLAND & G.L. ANDERSON, UNIV. LOUISVILLE

book, the reader must look beyond the imagery and illustrations, and read the numerous vignettes highlighted throughout the text, as well as the supplementary sidebars provided on the accompanying CD-ROM. Here Weinberg demonstrates the breadth of his understanding and insight, and discusses the problems still facing us. For example, if micrometastases are attracted to an environment where they can flourish, why are recurrences of renal cancer not centred on the other kidney? How do we overcome the elevated interstitial pressure inside tumours, which has bedevilled attempts to deliver therapeutics effectively? Furthermore, despite the promise of mechanism-based therapeutics and pharmacogenetics, the best predictor of survival after treatment of lung cancer with an anti-EGFR inhibitor is still smoking status.

There is no comparable text in cancer biology and no single book that is so current and

informative. Weinberg was ably assisted by an artist, Nigel Orme, in translating complex concepts into imaginative illustrations.

I would recommend the hardback copy: my soft-bound version came apart after one week of bed-time reading.

Although the book's stated intention is "to stimulate a new breed of cancer researchers", there is much here even for scholars to learn or be reminded of. Most molecular- and cell-biology students are familiar with *Molecular Biology of the Cell*, edited by Bruce Alberts and colleagues, now in its fourth edition (Garland Science, 2002). In a similar way, I hope that professionals and students in all spheres of cancer research will get to know *The Biology of Cancer*. ■

Chris Boshoff is professor of cancer medicine at the Wolfson Institute for Biomedical Research and the University College London Cancer Institute, Gower Street, London WC1E 6BT, UK.

Detecting hidden objects

Seeing Spatial Form

by Michael R. M. Jenkin & Laurence R. Harris

Oxford University Press: 2005. 449 pp.

£38.99

Manfred Fahle

Seeing Spatial Form covers an apparently eclectic mix of topics in visual perception, from 'pictorial relief' and 'vision in flying' to 'neuronal damage' and 'dopamine receptors'. In the introduction, the editors, Michael Jenkin and Laurence Harris, relate these and other seemingly disparate themes to the general topic of 'spatial form' — the detection of (often hidden) shapes in the visual world. In my view,

detecting objects' shapes in the enormous complexity of retinal signals is one of the most elaborate feats achieved by the human brain. One excellent example of this feat is the detection of camouflaged items, which is proving a tough problem for artificial intelligence.

The book does not provide a systematic treatment of the perception of spatial form in vision, so how did the editors decide what to include? They were guided by three principles: the choice of first-rate contributors; the adoption of a systems-analytical and model-driven approach; and the desire to select topics close to David Martin Regan's heart. This is because the book is a *Festschrift* in honour of Regan's 68th birthday. So, instead of an integrated,

textbook-like coverage of the perception of spatial form, the book provides rather specialized, mostly very well written, chapters that are more or less closely related to the topic, by some of the most prominent people in the field.

The individual chapters provide excellent reviews on state-of-the-art research in a wide range of topics, all related to Regan's interests. And wide they are, at a time when many researchers devote their life to a single topic, or even to a single technique. Regan is more of a Renaissance man, not only being interested in and using psychophysical methods, both in normal observers and patients suffering from different deficits of visual perception, but also sum-potential recordings, 'active' vision with eye movements, functional magnetic resonance imaging, neuropharmacology and developmental methods. Accordingly, the book covers a wide range of topics, and a wide range of methods for studying them.

It is not possible to mention all these topics

in a short review. Suffice it to say that Regan's friends and colleagues produced an excellent *Festschrift*, with well structured, well formulated and well argued synopses on many facets of visual perception, not just spatial form. Even a reader seasoned in vision research will find not just new pieces of knowledge, but plenty of new insights — this, in my opinion, makes it an ideal book to choose for those times when you're stranded on a desert island. It will tell you about all those topics in vision research that you always wanted to discover but did not have the time to read about.

Let me give you four examples. You probably know, or would have guessed, that attention plays an important role in programming saccades, or rapid eye movements. But were you aware that planning and executing them 'costs' cognitive resources, and that a batsman cannot follow an approaching ball with his eyes but needs to estimate its time of arrival with an accuracy better than one-hundredth of a second?

Second, it turns out that the processing of motion-defined form may be defective in patients even though simple motion detection is intact. Third, although retinal images are flat and the detection of surfaces represents an early stage of visual computation, perception is always inherently three-dimensional, even in the pictorial space of photographs. Finally, great progress has been made in understanding the neural basis of 'form vision'. For example, three-dimensional shape can be perceived from as few as two frames of a motion sequence.

If, once you have read the book, you want to know even more, no problem: there are selected references after each chapter, and a CD-ROM with additional material to keep you busy for quite a while.

Manfred Fahle is in the Department of Human Neurobiology, University of Bremen, 3 Argonnenstrasse, 28211 Bremen, Germany, and at the Henry Wellcome Laboratories for Vision Sciences, City University, London EC1V 0HB, UK.

Reaching for the stars

The beauty of a rocket engine designed by Valentin Glushko lies in its functionality.

Martin Kemp

How and when does a utilitarian piece of technology become an object of beauty? This question was triggered by an early rocket engine designed by Valentin Petrovich Glushko (1908–89) that is on display in the Museum of Space Exploration and Rocket Technology in St Petersburg.

Glushko's 'Rocket Motor 50' (shown here), elegantly sectioned to disclose the chamber where paraffin and nitric acid mixed, sits gleaming on a white plinth in a museum case. It was invented for a rocket conceived by Mikhail Tikhonravov, a visionary advocate of space travel. To eyes attuned to the aesthetic delights of high-tech objects by generations of twentieth-century artists, such as Russian sculptor Naum Gabo, this is a beautiful object.

Glushko, born in the Ukrainian city of Odessa, pioneered electric and liquid propulsion systems at the Gas Dynamics Laboratory, which then occupied the site of the present museum on Peter the Great's fortified island of Zayatchy. In 1935 he published his book *Rockets, their Construction and Utilization*.

Although imprisoned during a Stalinist purge, Glushko returned to become the leading engine designer for the Soviet missile and space programme. A falling-out with rocket designer Sergei Korolev, over the right propellants for the ill-fated Moon programme, left him sidelined again, but after Korolev's death he returned to head the space bureau.

It seems that Glushko's engine designs



were concerned solely with function. So is there even the slightest justification for seeing his rocket engines as anything other than obsolete relics of a great technological adventure?

There seem to be two reasons why our fascination with such relics goes beyond historical curiosity. The first is largely superficial: there is an undeniable romantic attraction to old technologies, once modern but now quaint. The crowds at shows of veteran cars or early aeroplanes testify to the emotional pull of antique machinery. An early electron microscope in a science museum exudes an air of nostalgic quaintness, even for spectators who know little of the history of such devices. Such

items have acquired a 'period style' that speaks eloquently of their eras.

The Glushko engine, however, has little period style, so the second factor must predominate. This is what I call the aesthetic of ineffable rightness. Supreme designs in technology and the applied arts often exude an intense air of inevitability. Once invented, we can see that the design presents an optimal solution that transcends the idiosyncrasies of its human designer. In relation to function and materials, we can intuit that the solution presented is the right one.

The form of Glushko's engine is dictated by technical parameters, such as the angles and bends of the pipes that feed fuel to the central cylinder. They were not designed to look good — but they do.

At a time when the United States is contemplating manned expeditions to Mars, those motivations that lie outside the strict scientific imperatives again come to the fore. The aesthetics of space flight — the beauty of the technology, whether sleek or quirky — are important if the programmes are to garner public support.

Was Glushko himself solely driven by scientific and technological imperatives? My guess, looking at his Rocket Motor 50, is that he was deeply engaged with the visual 'rightness' of his invention as an integral component in his quest for functionality.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.

NEWS & VIEWS

ARCHAEOLOGY

High tech from Ancient Greece

François Charette

The Antikythera Mechanism, salvaged 100 years ago from an ancient shipwreck, was long known to be some sort of mechanical calendar. But modern analysis is only now revealing just how sophisticated it was.

During renovation work in a northern Italian *palazzo*, an enigmatic artefact comes to light, dated to the late fifteenth century. After intensive analysis, it is identified as a complex steam engine — constructed 200 years before French inventor Denis Papin's pioneering experiments, and 300 years before the Industrial Revolution. Our view of the technical achievements of the Renaissance is completely changed. The reverberations are felt far beyond just scholarly circles.

True, this hasn't happened. But a century-old archaeological find is continuing to force a comparable rethink of the technology of classical antiquity. In this issue, Freeth *et al.* (page 587)¹ present the most up-to-date analysis of this artefact, the Antikythera Mechanism.

In 1900, a team of Greek sponge-divers working off the islet of Antikythera, midway between the Peloponnese and Crete, discovered an antique shipwreck 42 metres below the surface of the Mediterranean Sea. Among the many objects they recovered from the wreck, which has been dated to around 65 BC, were several bronze fragments. At first overlooked, these were later associated with some sort of astronomical machinery. But the realization that this was the earliest-known device involving an arrangement of gear-wheels came only slowly. In fact, staggeringly, the Antikythera Mechanism is the most sophisticated such object yet found from the ancient and medieval periods.

From the late 1950s to the early 1970s, the late historian of science and technology Derek De Solla Price studied the badly corroded and fragmented mechanism extensively, especially once γ -radiography techniques had become available. This work culminated in a lengthy essay, *Gears from the Greeks*². In it, Price proposed that the mechanism had been an astronomical calendar with display dials on the front and back indicating the positions of the Sun and the Moon. This gearing, comprising some 30 wheels, was, according to Price, a clever mechanical simulation of a basic period relation that linked the length of the solar year with the phases of the Moon. This relation is the Metonic cycle, which says that in 19 solar

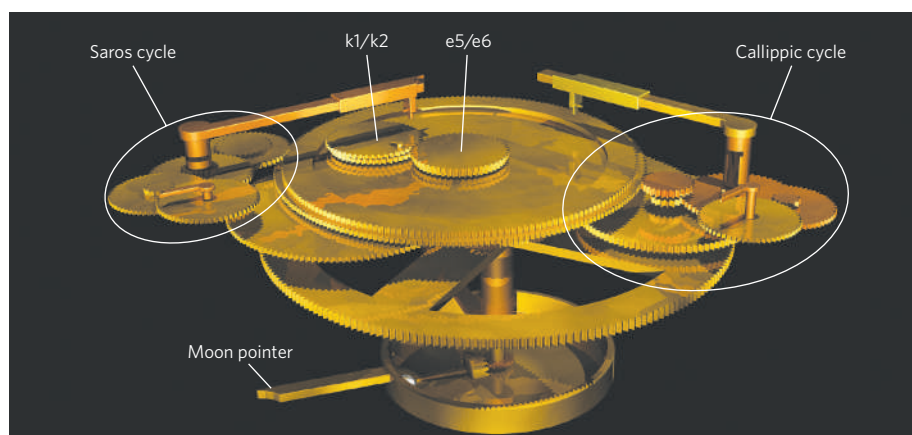


Figure 1 | Wheels within wheels. The rear side of Freeth and colleagues' reconstruction¹ of the Antikythera Mechanism, viewed sideways on. The left gear and pointer system simulated the Saros cycle for predicting lunar and solar eclipses; the right gears and pointers were for the Callippic cycle that synchronizes synodic months and solar years. At the centre, mounted on the large gear-wheel, were two pairs of identical gear-wheels, e5/e6 at the centre and k1/k2 at the left (see also Fig. 5 on page 590). The pair k1/k2 was provided with a pin-and-slot device that induced an irregular movement in the pointer at the front of the mechanism indicating the position of the Moon. This system simulated a model of the Moon's motion developed by Hipparchus of Rhodes in the second century BC.

years there are 235 lunations (the time from new Moon to new Moon, also known as a synodic month) and 254 sidereal months (the time the Moon takes to reach the same location with respect to the fixed stars).

Price's heroic analysis and reconstruction, and his claim that the mechanism "requires us to completely rethink our attitudes toward ancient Greek technology"³, and perhaps even our understanding of Hellenistic and Graeco-Roman civilization, did not meet with the enthusiasm, or even generate the widespread recognition, that it perhaps deserved. Since the late 1980s, various criticisms have been levelled at the reconstruction, and a number of alternative proposals have been made. Price, it seems, was a little too quick to adopt tooth numbers and arrangements that satisfied his preconceived idea of the nature and function of the mechanism.

Great advances in deciphering the Antikythera Mechanism have been made in recent years by Michael Wright, who has relied on digitized X-rays and linear tomography to achieve better imaging of the mechanism and

some remarkable insights into its functions⁴. Freeth and colleagues' paper¹ now details the investigations of a multidisciplinary team from Britain, Greece and the United States that reanalysed the Antikythera Mechanism using the best available technology — three-dimensional X-ray computer tomography and high-resolution surface imaging.

Compared with Price's readings, twice as many textual data are now legible from the Greek inscriptions on the mechanism. Analysis of the now much clearer lettering indicates a construction date of 150–100 BC, slightly earlier than had been assumed. Freeth and colleagues clarify the function of the front and back dials of the mechanism: on the front were graduations for the zodiac and the solar calendar, and pointers for the Sun and Moon with an indication of the lunar phase. The back dials indicated time in terms of two astronomical cycles. Each dial involved a pointer with an ingenious spiral design (Fig. 1). One dial is for the Callippic cycle. This cycle is a later attempt to improve the Metonic cycle's prediction of the relation between lunar months and solar

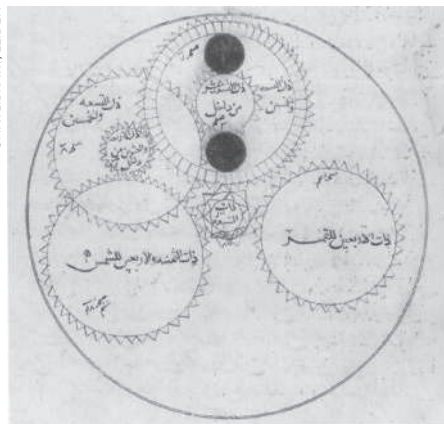


Figure 2 | Islamic descendant. An eight-gear lunisolar calendar illustrated in al-Biruni's treatise on the astrolabe, written in AD 996. The design is much simpler than that of the Antikythera Mechanism, but is very probably descended from it.

years, and holds that 940 lunations occur in 76 years (four Metonic cycles) less one day. A second dial is for the Saros cycle of 223 lunations (roughly 18 years), which was used to predict the occurrence of solar and lunar eclipses.

Freeth and colleagues' reconstruction¹ involves 37 gear-wheels, of which seven are hypothetical; Price found 29 gears and proposed a further two². In the face of fragmentary material evidence, such guesswork is inevitable. But the new model is highly seductive, and convincing in all of its details. It ought to force us, definitively, to abandon Price's reconstruction, which is still frequently reproduced in general and scholarly books.

Among the most important of the authors' conclusions¹ is that a highly ingenious pin-and-slot device connecting two superimposed gear-wheels, one slightly off-centre, induced a quasi-sinusoidal variation in the movement of the Moon in the mechanism (Fig. 1). They show this to be a mechanical realization of a geometrical model of the first lunar anomaly — an initial approximation of the irregular nature of the Moon's motion — that had been developed by the astronomer Hipparchus of Rhodes in the second century BC. The authors speculate that Hipparchus might even have been involved in the initial design of the mechanism; there is circumstantial evidence that the ship that carried the Antikythera Mechanism set sail from Rhodes, based principally on the Rhodian origins of artefacts such as amphorae and coins found on it². Newly deciphered inscriptions that relate to the planetary movements make it plausible that the mechanism originally also had gearings to predict the motions of the planets.

Is this the last word on the Antikythera Mechanism? Certainly not, but it does provide a new standard, and a wealth of fresh data, for future research. In the meantime, although stunned at the ingenuity of the ancients, we will still be tempted to continue speculation about the specific functions fulfilled by the

mechanism. Beyond that, there is the more substantive question of the place and scope of such advanced technology in the ancient Eastern Mediterranean world, and of its eventual fate. It was a long time before gearing mechanisms of this sophistication re-emerged, at least on the current archaeological record. Certain elements of the mechanism are encountered, albeit in much simpler design, in fifth-century Byzantium, and again in medieval Islam^{2,5}. The celebrated Persian scholar al-Biruni described, shortly before AD 1,000, a mechanical lunisolar calendar that was usually inserted within an astrolabe with displays on the back, and that approximated the Metonic cycle by means of eight gear-wheels (Fig. 2)⁶. An astrolabe from thirteenth-century Iran containing such a device is in the Museum of the History of Science at the University of Oxford, UK.

Al-Biruni mentioned two alternative gearing arrangements for such a mechanical calendar commonly used by contemporary craftsmen. An anonymous Arabic treatise has recently surfaced that precisely describes one of them, based on a purely lunar calendar. The mechanism is attributable to Nastulus, an

instrument-maker and astronomer active in Baghdad around AD 900. This tradition probably derives from ancient technology that was either an antecedent form or a later simplification of the Antikythera Mechanism. But it is equally obvious that much of the mind-boggling technological sophistication available in some parts of the Hellenistic and Graeco-Roman world was simply not transmitted further. The gear-wheel had, in this case, to be reinvented. The Antikythera Mechanism is a useful reminder that history seldom follows simple, linear paths. ■

François Charette is at the Lehrstuhl für Geschichte der Naturwissenschaften, Ludwig-Maximilians-Universität, Museumsinsel 1, D-80538 München, Germany. e-mail: fcharette@ankabut.net

1. Freeth, T. et al. *Nature* **444**, 587–591 (2006).
2. Price, D. de S. *Trans. Am. Phil. Soc.* **64**(7), 1–70 (1974).
3. Price, D. de S. *Science Since Babylon* 48 (Yale Univ. Press, 1975).
4. Wright, M. T. in *Proc. 2nd Int. Conf. Ancient Greek Technol.* (ed. Tassios, Th.) 49–60 (Technical Chamber of Greece, Athens, 2006).
5. Wright, M. & Fields, J. *Ann. Sci.* **42**, 87–138 (1985).
6. Hill, D. *Ann. Sci.* **42**, 139–163 (1985).

STEM-CELL BIOLOGY

A move in the right direction

Jeffrey S. Chamberlain

Stem-cell therapy is valued for its potential to restore damaged or degenerating tissues. Stem cells are now regularly used to renew blood, and it looks as if the next success could be in treating dystrophic muscle.

The potential of stem-cell technologies to revolutionize medical care is causing great excitement among biologists and the general public. Recent studies on embryonic and adult stem cells, coupled with advances in our understanding of how they can be coaxed into forming particular cell types and tissues, have improved the prospects for addressing a host of untreatable diseases. Bone-marrow transplants to renew blood stem cells are now routine, but significant technological hurdles must still be overcome before the power of stem cells can be harnessed to regenerate solid organs such as muscle. On page 574 of this issue, Giulio Cossu and colleagues (Sampaioles et al.)¹ describe evidence from an animal model that a fairly straightforward infusion of stem cells into the bloodstream might one day be able to treat muscular dystrophy*.

Cossu and colleagues focused on Duchenne muscular dystrophy, one of the most common and devastating inherited disorders, which is caused by mutations in the gene that encodes the dystrophin protein. A treatment for this

condition has long been sought², and a flurry of breakthroughs has led to current or planned clinical trials of at least five different treatment strategies³. Prominent among these are stem-cell transplantation and gene therapy, two approaches that can also be combined by manipulating an individual's own (autologous) stem cells to 'correct' the genetic defect before transplantation. Cossu and colleagues studied an animal model of Duchenne muscular dystrophy — the dystrophic GRMD dog — that closely mimics the human disease, in particular showing extensive muscle wasting and early death (unlike mouse models of the disease). They transplanted either donor stem cells or corrected autologous stem cells into the dogs and observed dramatic improvements in the animals' muscles.

A key factor in this work was the previous identification by the Cossu lab of a novel type of adult stem cell, termed a mesoangioblast, which can be harvested from small blood vessels⁴. Like all stem cells, mesoangioblasts can divide to make more of themselves, but they are also 'preprogrammed' to develop into muscle cells. Although a variety of stem-cell types can become muscle cells when injected into the

*This article and the paper concerned¹ were published online on 15 November 2006.

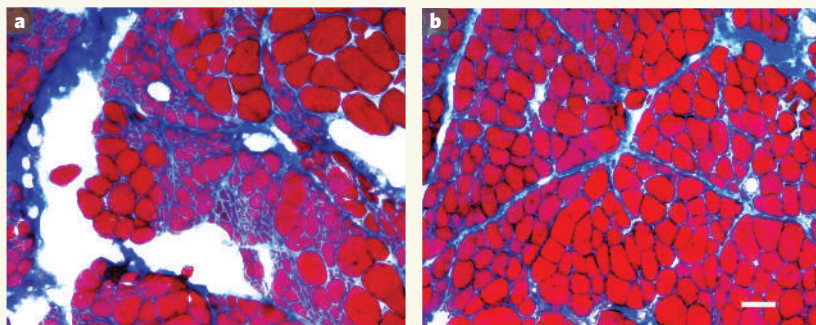


Figure 1 | Stem-cell treatment of dystrophic muscle in dogs¹. These pictures are stained biopsies of (a) untreated dystrophic skeletal muscle and (b) muscle after treatment. The latter was both structurally more ordered and functioned better. Scale bar, 200 μ m.

bloodstream or directly into muscles^{5,6}, mesoangioblasts show several distinct advantages over these other cells. They are relatively easy to isolate, and their numbers can be expanded greatly in tissue culture without losing the ability to form muscle. When mesoangioblasts are infused into arteries, they pass through vessel walls to engraft within, and rescue, damaged muscle cells with a surprisingly high efficiency⁷. Such properties are ideal for treating a disease such as muscular dystrophy in which muscles all over the body need to be repaired.

In the present study¹, Cossu and colleagues isolated mesoangioblasts from normal and dystrophic dogs, expanded them in culture and infused them into a major hindlimb artery of the dystrophic dogs. Before infusion, the dystrophic mesoangioblasts were genetically 'corrected' by inserting copies of a gene that encodes a micro-dystrophin — a truncated form of the dystrophin protein — along with all the instructions needed to make the protein in muscle. These shortened dystrophins were used because they correct most dystrophic abnormalities in mouse models and can be transferred into stem cells far more efficiently than the full-length dystrophin⁸.

Infusion of either the normal or corrected dystrophic mesoangioblasts led to new expression of the dystrophin protein in a variety of different muscle groups in the dystrophic dogs. Dystrophin was also found in some muscles from the side of the dogs' bodies that had not been injected, indicating that mesoangioblasts could travel great distances before engrafting. To improve the efficiency of dystrophin production, the dogs were injected with cells up to five times at monthly intervals. In one animal, the cells were released from a catheter into the aorta — the vessel leaving the heart that supplies the body with blood. This therefore allowed more widespread dissemination of the mesoangioblasts. The results of the stem-cell infusions were dramatic: this last animal displayed a marked improvement in its dystrophy and was walking well 5 months after the final injection; the other animals recovered to a lesser degree. In general, dogs receiving donor cells improved more than those receiving corrected autologous mesoangioblasts.

Some muscles in the injected dogs had nearly normal levels of dystrophin, and even the muscles with only moderate levels of dystrophin showed significantly improved structure and function (Fig. 1).

The different results obtained with normal and autologous mesoangioblasts have implications for their application in human patients. As with any transplant, unless the donor and recipient are immunologically matched, the recipient must be placed on lifelong immune suppression; this is not always effective and can have nasty side effects. The use of genetically corrected autologous stem cells would avoid this necessity, and so would be preferable if its effectiveness could be improved. There are at least two possible reasons for the reduced effectiveness of the autologous stem cells compared with the donor ones. First, micro-dystrophins are not fully functional, so alternative mini- or micro-dystrophins with improved functional activity⁹ might be more effective. Second, the

micro-dystrophin used to correct the dystrophic mesoangioblasts was a human protein, and better results might have been obtained with dog dystrophin.

Cossu and colleagues' results¹ provide compelling evidence that this method should be developed further for testing in patients. That will take several years, and may need many rounds of refinement in animals before any human trials can take place. Importantly, the corrected mesoangioblast approach could have widespread application to a variety of other muscle diseases besides Duchenne muscular dystrophy. The Cossu lab has shown that mesoangioblasts can improve muscle function in a mouse model of another type of muscular dystrophy¹⁰. There are more than 20 types of muscular dystrophy and numerous other muscle disorders, but treatment options are almost non-existent^{2,3}. Perhaps these conditions will be among the first for which the promise of stem-cell technology for degenerative disorders can be realized. ■ Jeffrey S. Chamberlain is in the Department of Neurology, University of Washington, Seattle, Washington 98195-7720, USA. e-mail: jsc5@u.washington.edu

1. Sampaioles, M. et al. *Nature* **444**, 574–579 (2006).
2. Emery, A. E. *Lancet* **359**, 687–695 (2002).
3. Chamberlain, J. S. & Rando, T. A. (eds) *Duchenne Muscular Dystrophy: Advances in Therapeutics* (Taylor & Francis, New York, 2006).
4. Minasi, M. G. et al. *Development* **129**, 2773–2783 (2002).
5. Gussoni, E. et al. *Nature* **401**, 390–394 (1999).
6. Qu-Petersen, Z. et al. *J. Cell Biol.* **157**, 851–864 (2002).
7. Galvez, B. G. et al. *J. Cell Biol.* **174**, 231–243 (2006).
8. Gregorevic, P. et al. *Nature Med.* **12**, 787–789 (2006).
9. Harper, S. Q. et al. *Nature Med.* **8**, 253–261 (2002).
10. Sampaioles, M. et al. *Science* **301**, 487–492 (2003).

CHEMICAL BIOLOGY

A broader take on DNA

Aaron M. Leconte and Floyd E. Romesberg

Slipping in extra benzene rings creates a broader DNA double helix that is similar to, but different from, natural DNA. Importantly, it can encode more genetic information — and that could have wide implications.

Compared with proteins, which can have any number of jumbled forms, the DNA double helix — that most famous of nature's molecular structures¹ — is comfortably regular. Over the past few years, however, Eric Kool and his colleagues have re-engineered this predictable duplex, producing an 'expanded' version that they call xDNA². Writing in the *Journal of the American Chemical Society*³, these researchers further elucidate the structure of this architectural delight. Their results indicate that, like a highway with a lane added to convey more traffic, the wider duplex can carry more genetic information than can standard DNA.

Probing and manipulating the dimensions of nucleic acids is nothing new for chemists. In fact, the first synthesis and characterization of size-expanded DNA bases was carried out in the 1970s⁴, and expanded bases are still being developed and used to study protein specificity, and as a route to drug development^{5,6}. But these studies all concentrate on the single basic unit of DNA — the nucleotide — rather than the oligonucleotide polymer from which DNA strands are made. Full-blown, double-stranded DNA is constructed by the selective pairing of the bases of two oligonucleotides through hydrogen bonding to form the 'rungs' of

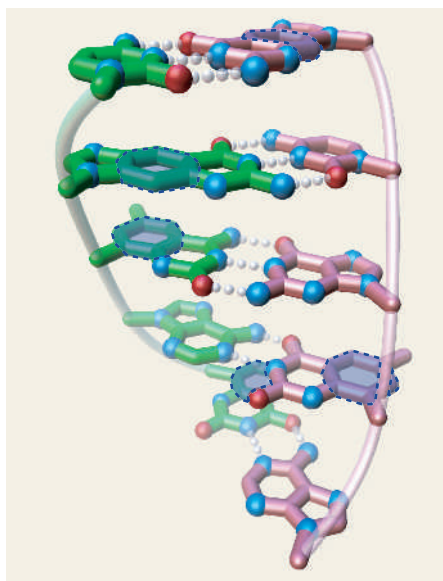


Figure 1 | Expanded DNA. In xDNA, depicted above, a normal base and an expanded base (with an extra benzene ring, indicated by blue dotted lines) pair up according to hydrogen bonding and size complementarity to form a duplex with four possible base pairs. (Figure courtesy M. Pique, Scripps Res. Inst.)

the duplex. Kool and colleagues' study is unique in that it exploits the expanded bases to create new criteria for base-pairing among oligonucleotides, and thus produces an entirely different duplex structure.

xDNA is made up of analogues of the nucleotides of natural DNA — guanine (G), cytosine (C), adenine (A) and thymine (T) — that are dubbed xG, xC, xA and xT. These nucleotides carry bases that are identical to their natural counterparts, but they are expanded at right angles to the duplex's long axis by the addition of a benzene ring (Fig. 1). In natural DNA, G always pairs up with C, and A with T. The expanded analogues maintain their natural hydrogen-bonding topology, so they can match up with their complementary base to form pairs that are wider than normal by a single benzene ring. But accommodating a single expanded pair within a natural duplex requires significant distortion of the duplex backbone, and is therefore energetically disfavoured. Likewise, a pair formed between two natural bases is too narrow to fit into a duplex in which all the other pairs are expanded.

Previous studies by Kool and colleagues^{2,7} have shown that two strands of modified DNA can combine to form a stable xDNA duplex provided that just one member of every base pair is expanded. This is true regardless of whether the expanded bases are all within one strand, or spread between both. Given the hydrogen-bonding requirements, there are thus four viable partnerships in the expanded duplex: xG–C, G–xC, xA–T and A–xT. This is twice as many as in natural DNA, so the expanded duplex can encode more information.

In their latest paper³, Kool *et al.* present us with the first picture of xDNA containing all

four possible base pairs. The studies confirm that such an xDNA duplex bears many of the same features as B-DNA (the right-handed-twisting variant that is the most common natural form of DNA). But it has several interesting differences, too. For example, the new structure reveals extensive inter- and intra-strand stacking between expanded bases. This stacking probably explains the remarkable stability of the xDNA duplex. And, to accommodate the increased diameter of the bases, twelve base pairs are required to complete one turn of the expanded duplex, rather than the ten of the natural duplex. The so-called major and minor grooves of the xDNA duplex are therefore also markedly wider and shallower. As these grooves mediate many of the interactions that underlie protein and small-molecule binding to DNA, such molecules might interact with xDNA in different, and perhaps useful, ways.

The spontaneous assembly of xDNA into a unique and regular duplex containing four unique base pairs could make the material useful in various bio- and nanotechnological applications. Unlike natural DNA, xDNA bases are fluorescent, making spectroscopy easier to perform. And as xDNA can form twice as many base pairs as natural DNA, it could also provide an unexplored avenue towards expanding the genetic alphabet⁸.

Daunting challenges remain. Most notably, taking advantage of the increased information potential of xDNA will require, at a minimum, a polymerase capable of replicating the duplex. The polymerases that replicate natural DNA are very sensitive to the geometry of a base pair⁹, so it seems unlikely that any natural polymerase will accommodate the altered dimensions of the xDNA. Although the directed evolution of polymerases — creating variants of the enzyme

with specific functions through *in vitro* mutagenesis and selection — has had some success¹⁰, something as radical as producing an xDNA polymerase has never been attempted. This would in itself be a milestone in directed evolution.

The structure of xDNA raises as many fascinating questions as it answers. It will be particularly interesting to see what kinds of duplex structures DNA forms when its bases are expanded in different directions. Kool and colleagues have already reported several other types of expanded base¹¹, so answers could be swift in coming. Further biochemical and biophysical study of these duplexes should uncover more surprises and also provide general information about interactions between nucleic acids and proteins. It is no small accomplishment to re-engineer one of nature's oldest and most universal molecules. Like its slimmer progenitor, xDNA should be around for years to explore, study and enjoy. ■

Aaron M. Leconte and Floyd E. Romesberg are in the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: floyd@scripps.edu

1. Watson, J. D. & Crick, F. H. *Nature* **171**, 737–738 (1953).
2. Liu, H. *et al.* *Science* **302**, 868–871 (2003).
3. Lynch, S. R., Liu, H., Gao, J. & Kool, E. T. *J. Am. Chem. Soc.* **128**, 14704–14711 (2006).
4. Leonard, N. J. *Acc. Chem. Res.* **15**, 128–135 (1982).
5. Seley, K. L., Quirk, S., Salim, S., Zhang, L. & Hagos, A. *Bioorg. Med. Chem. Lett.* **13**, 1985–1988 (2003).
6. Zhang, N. *et al.* *J. Med. Chem.* **46**, 4776–4789 (2003).
7. Gao, J., Liu, H. & Kool, E. T. *Angew. Chem. Int. Edn* **44**, 3118–3122 (2005).
8. Henry, A. A. & Romesberg, F. E. *Curr. Opin. Chem. Biol.* **7**, 727–733 (2003).
9. Johnson, S. J. & Beese, L. S. *Cell* **116**, 803–816 (2004).
10. Henry, A. A. & Romesberg, F. E. *Curr. Opin. Biotechnol.* **16**, 370–377 (2005).
11. Lee, A. H. & Kool, E. T. *J. Am. Chem. Soc.* **127**, 3332–3338 (2005).

CONSERVATION BIOLOGY

Rarity bites

Barry W. Brook and Navjot S. Sodhi

Rare species have to cope not only with habitat loss, genetic bottlenecks and invasive competitors, but also with a self-reinforcing cycle of human greed. This last threat has now been dragged into the spotlight.

It makes sense that, as a valued commodity becomes scarce, its cost rises. Indeed, tests of the economic theories of demand have shown that consumers greatly value the hedonic exclusivity of owning rare objects, such as coins or stamps¹. But is this desire restricted to the harmless venture of collecting historical curiosities? Writing in *PLoS Biology*, Courchamp and colleagues² suggest not. They describe an under-appreciated thorn in the side of endangered species already teetering on the brink of extinction — the lure of the few.

Standard bioeconomics asserts that wildlife harvests will halt when the expense of locating individuals of a species exceeds the financial gain, thus shielding that species from fatal overexploitation. However, Courchamp *et al.* develop a mathematical model to show that the exploitation of rare and endangered species can have a paradoxical outcome. If rarity makes them more desirable to some, then ever-increasing pecuniary incentives for poaching can create a positive feedback loop, ending in the species' demise. The authors show how the



50 YEARS AGO

On September 18 the Commonwealth Scientific and Industrial Research Organization of Australia held a conference in Melbourne to plan the country's myxomatosis campaign for the season... Reports from all the Australian states indicated that rabbit populations are extremely low. Myxomatosis has been largely responsible for the reduction in rabbit numbers, but other factors have also contributed... The conference urged all landholders to use poison '1080' and all other effective methods after the current myxomatosis season ends. Advantage should be taken of low rabbit numbers to kill off all survivors. The aim must be to eradicate the rabbit from Australia.

From *Nature* 1 December 1956.

100 YEARS AGO

Lieut.-Colonel C. D. Durnford has a second paper on the flying-fish problem... [T]he author in his original paper...endeavoured to prove on mathematical grounds that the "aëroplane theory" of the flight of these fishes was a physical impossibility, owing to the relatively small wing-surfaces, and that consequently progression through the air must be due to intensely rapid wing-vibration, aided in certain circumstances by movements of the tail... In the supplementary communication Colonel Durnford adduces further evidence in favour of this explanation of the phenomenon... [T]he chief features of the flight appear to be as follows:— (1) the tail-impelled, visibly wing-assisted jump from the water to a height where the wings can work visibly; (2) the flight continued by an intensely rapid and laboured wing-movement, generally mistaken for a condition of rest, and, if seen at all, visible only as a blur; (3) short periods of slowing down of wing-movement, when the vibrations again become perceptible; (4) either sudden cessation of wing-movement...or a short slow-down into visibility immediately preceding immersion.

From *Nature* 29 November 1906.

T. TAN



Figure 1 | Object of desire — the critically endangered Bali starling. This species has been heavily collected for the pet trade, and is now on the brink of extinction in the wild.

value placed on rare species for collections, as luxury items, as pets or trophies, in traditional medicine, and even in ecotourism, can keep pace with the cost of acquiring such species.

Courchamp *et al.* support their case by providing a series of telling examples covering plants, molluscs, insects, fish, amphibians, reptiles, birds and mammals. For instance, the swim bladder of the Chinese bahaba (*Bahaba taipingensis*) is so valued for traditional medicine that its price per unit weight eventually exceeded that of gold by sevenfold. Even when fewer than half-a-dozen fish were being caught each year in the 1990s, more than 100 boats continued to try to catch them². There are many other examples^{3–5}, illustrating how widespread this problem may be. The Bali starling (*Leucopsar rothschildi*; Fig. 1) epitomizes how the craving for rare birds in the pet trade can drive a species to the edge of extinction. Despite being listed in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)⁶ since 1970, 19 individuals were reported as being sold illegally in shops in Singapore in 1979, and 16 were observed in cages in Denpasar (Bali) in 1982. This was when fewer than 150 birds were left in the wild. They now number six, restricted to Bali Barat National Park⁵.

Courchamp *et al.*² call this rarity-value feedback the 'anthropogenic Allee effect'. Allee effects refer to phenomena that strongly influence the dynamics of populations only at low densities, and they include the increasing difficulties of finding mates, of maintaining a cohesive social structure, and of avoiding the genetic hazards associated with inbreeding. By reducing the chances of successful reproduction, Allee effects are synergistic and self-reinforcing, and can plunge rare species into what has been colourfully described as an 'extinction vortex'⁷. An empirical analysis of the final chapter in the life story of some well-

studied species suggests that the extinction vortex is a real problem⁸.

Courchamp and colleagues' model requires that consumers are always willing to pay what it takes to obtain a rare species, making extinction by collection an inevitable prediction. Yet all that may be required is for over-harvesting to continue long enough to push rare species below a population size that dooms them to an extinction vortex. If so, extinction induced by the anthropogenic Allee effect might be a disturbingly common event, even if collecting ceases before the species' final demise. This may have happened in the case of the Tasmanian tiger (*Thylacinus cynocephalus*), whose population crashed during a period of bounty hunting, but which then dwindled to extinction after exploitation had ended⁹.

It is essential to identify threatened species and populations so that action can be taken to protect them. The best-known system of this kind is the IUCN Red List drawn up by the World Conservation Union¹⁰. Publicizing the plight of vulnerable plants and animals is crucial because it helps to secure public support and funds. But in so doing, are conservationists unintentionally making things worse? It seems that the very act of identifying new rare species, and describing where they are, can lure collectors and so induce an anthropogenic Allee effect, with sometimes devastating consequences¹¹. Perhaps controversially, Courchamp *et al.*² also provide evidence that declaring a species' threatened status on the IUCN Red List and CITES Appendices amplifies the black-market demand for that species.

Breaking this vicious circle will be difficult and potentially costly, but necessary. The possibilities include keeping secret the location of newly designated rare species; commercial rearing and subsequent legal sale of otherwise rare species or their products, such as traditional medicines; and enforcing measures that make poaching too risky, as is the case in some African game reserves where armed guards shoot to kill. Of course, the most effective (but idealistic) strategy would be to curb demand. As Mahatma Gandhi famously said: "Earth provides enough to satisfy every man's need, but not every man's greed."

Barry W. Brook is in the School for Environmental Research, Charles Darwin University, Darwin, Northern Territory 0909, Australia. Navjot S. Sodhi is in the Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543 Singapore.
e-mails: barry.brook@cdu.edu.au;
dbsns@nus.edu.sg

1. Koford, K. & Tsochoglou, A. E. *J. Econ. Behav. Organ.* **34**, 445–457 (1998).
2. Courchamp, F. *et al.* *PLoS Biol.* **4**, e415 (2006).
3. Rodríguez-Estrella, R., Carmen, M. & Moreno, B. *Biodivers. Conserv.* **15**, 1621–1625 (2006).
4. Webb, J. K., Brook, B. W. & Shine, R. *Oryx* **36**, 170–181 (2002).
5. Sodhi, N. S. & Brook, B. W. *Southeast Asian Biodiversity in Crisis* (Cambridge Univ. Press, 2006).

6. www.cites.org
7. Gilpin, M. E. & Soule, M. E. in *Conservation Biology: The Science of Scarcity and Diversity* (ed. Soule, M. E.) 19–34 (Sinauer, Sunderland, MA, 1986).
8. Fagan, W. F. & Holmes, E. E. *Ecol. Lett.* **9**, 51–60 (2006).

9. Bulte, E. H., Horan, R. D. & Shogren, J. F. *Ecol. Econ.* **45**, 271–279 (2003).
10. www.redlist.org
11. Stuart, B. L., Rhodin, A. G. J., Grismer, L. L. & Hansel, T. *Science* **312**, 1137 (2006).

ORGANIC CHEMISTRY

Molecules in quarantine

Julius Rebek Jr

Intermediate compounds are often produced during a chemical reaction, but they are too short-lived to be easily observed. It seems that a molecular pyramid can persuade them to stick around for a little longer.

The path from starting materials to products in chemical reactions often involves many steps. Intermediate compounds are the resting stages along this path, but they exist in vanishingly small concentrations and can rarely be observed. The methods used to see them all require isolation of the intermediates — a sort of quarantine that keeps them from encountering other molecules and so prevents them from continuing along the reaction pathway.

Several methods have been developed to enforce this isolation. For example, reaction intermediates in the gas phase at very low pressures can be kept apart long enough for them to be observed by spectroscopy. Alternatively, exotic solvents such as liquid helium droplets or frozen argon matrices can isolate intermediates at temperatures near absolute zero. But the segregation of reaction intermediates at room temperature, in a reactive solvent such as water, is far more difficult. That is just what Dong *et al.*¹ have done, as they report in the *Journal of the American Chemical Society*.

The authors isolated intermediates for a widely used type of reaction that is catalysed by amine compounds. They did this by surrounding the intermediates with a capsule that temporarily protects them from further reaction. The capsule is a tetrahedron-shaped complex of four gallium ions (at the corners) and six organic molecules (along the edges; Fig. 1). This forms spontaneously from its constituent parts and is stable in water; its overall negative charge provides an ideal nanoenvironment for positively charged molecules that can fit inside.

The intermediates isolated by Dong *et al.* are positively charged iminium ions, which form when amines react with ketone molecules (Fig. 1). The concentration of these intermediates in water is tiny, as they react rapidly with water to regenerate the starting materials. The authors¹ found that merely mixing the reaction components — capsule, amine and ketone — in water was enough to generate encapsulated iminium ions in good yields. The yields varied depending on the molecular size of the amines and ketones used, but typically 30–90% of the capsules were occupied by iminium ions. The trapped intermediates were detected directly

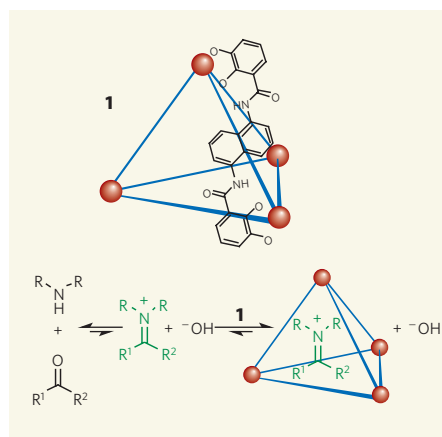


Figure 1 | Trapping reactive intermediates.

Molecular tetrahedra (**1**) form spontaneously in water from four gallium ions (red) and six organic molecules (only one of these molecules is shown for simplicity; their positions are represented as blue lines). Dong *et al.*¹ used these tetrahedra to trap iminium ions (green) that form fleetingly in water from their starting materials (R, R¹ and R² represent hydrocarbon groups).

by nuclear magnetic resonance (NMR) spectroscopy. The magnetic environment inside the capsule is unusual because of the many aromatic rings that enclose it. Encapsulated molecules thus show profoundly altered NMR signals compared with free molecules in solution, and so are easily distinguished from their unrestrained counterparts.

Given a choice of competing iminium ions with different sizes and shapes, the capsule preferentially traps the ion with the best fit, in a process known as molecular recognition. But although such competition experiments¹ go some way to defining the space inside the capsule, questions still remain. Given the flexibility of the larger iminium ions and the sizeable holes in the capsule's structure, it is impossible to say conclusively how much space there is and what fraction of it is occupied. Moreover, not all iminium ions are stabilized by this capsule, with the shape and size of the amine starting material being particularly limiting.

The same researchers have previously used

their capsule as a catalyst for a different kind of reaction². In this case, the captured starting material (an enammonium ion) underwent a molecular rearrangement, generating iminium ions. These particular ions were not held tightly by the capsule and so escaped, whereupon they were cleaved by the water solvent to yield two products. This overall process shows some of the essential features of enzymatic catalysis, such as selective recognition of starting materials and rapid product release. Indeed, there are many parallels between the encapsulation technique and enzymatic processes: reversible encapsulation can recognize, amplify³ and stabilize⁴ reactive intermediates and accelerate reactions under mild equilibrium conditions, just as enzyme active sites do. Product release from capsules is often difficult, but even that problem has been overcome in specific reactions⁵. And, like enzymes and other natural receptors, synthetic cavities have been fashioned that create an asymmetric environment; this should allow the recognition and control of molecular asymmetry in reactions, which is essential for preparing compounds that have biological activity.

The encapsulation technique also has implications for nanoscience. Capsules with huge cavities⁶, capable of holding eight molecules, have been analysed, and polymeric capsules that self-assemble⁷ have been discovered. These represent new ways of arranging molecules and might find applications in information storage. The ability to organize molecules within molecules within molecules is also an exciting prospect for researchers seeking to build materials from the inside out.

The reaction of amines with ketones is one of the most useful transformations in an organic chemist's repertoire, so Dong and colleagues' isolation of the elusive iminium intermediates is big news. The authors hope to design and prepare different capsules, so that other fleeting reaction intermediates can be stabilized and studied. It may even be possible to develop chemical reactions that were previously unthinkable in aqueous solution. The future of encapsulation chemistry seems assured. ■

Julius Rebek Jr is at the Skaggs Institute for Chemical Biology and in the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

e-mail: jrebek@scripps.edu

1. Dong, V. M., Fiedler, D., Carl, B., Bergman, R. G. & Raymond, K. N. *J. Am. Chem. Soc.* **128**, 14464–14465 (2006).
2. Fiedler, D., Bergman, R. G. & Raymond, K. N. *Angew. Chem. Int. Edn* **43**, 6748–6751 (2004).
3. Iwasawa, T., Mann, E. & Rebek, J. Jr *J. Am. Chem. Soc.* **128**, 9308–9309 (2006).
4. Yoshizawa, M., Kuskawa, T., Fujita, M. & Yamaguchi, K. *J. Am. Chem. Soc.* **122**, 6311–6312 (2000).
5. Yoshizawa, M., Tamura, M. & Fujita, M. *Science* **312**, 251–254 (2006).
6. MacGillivray, L. R. & Atwood, J. L. *Nature* **389**, 469–472 (1997).
7. Hof, F., Craig, S. L., Nuckolls, J. & Rebek, J. Jr *Angew. Chem. Int. Edn* **41**, 1488–1508 (2002).

NEUROSCIENCE

A memory boost while you sleep

Robert Stickgold

It is generally agreed that sleep aids memory consolidation, but the reasons for this are a mystery. Part of the answer may lie in the patterns of synchronous brain activity unique to the state of slumber.

Only ten years ago, discussions about the purpose of sleep offered great hypotheses, but these were based on flimsy evidence. So scant were the data that some researchers argued that sleep might have no use at all. This led Alan Rechtschaffen, a pioneer in the area, to comment wryly that “if sleep doesn’t serve an absolutely vital function, it is the biggest mistake evolution ever made.” Since then, researchers have produced a wealth of evidence for at least one function of slumber — the consolidation of memories. A wide range of converging data show that memories are replayed¹, modified², stabilized³ and even enhanced⁴ as we snooze, but an understanding of how this occurs remains elusive. The proposed mechanisms rely on two effects that are observed in the brain only during sleep: alterations in the levels of chemical neuromodulators⁵ and distinctive oscillations of electrical activity⁶. In this issue (page 610)⁶, Marshall *et al.* provide evidence for the second mechanism*. They show that the direct induction of ‘cortical slow oscillations’ during sleep can improve the recall of word-pairs memorized the previous night.

Human sleep is divided into rapid-eye-movement sleep (REM) and non-REM sleep (NREM), with NREM sleep further divided into stages 1 to 4. These sleep phases are distinguished, in part, by well-defined patterns of oscillatory electrical activity in the brain, as measured by electroencephalography. Theta waves oscillate at 4–8 Hz, and are characteristic of REM sleep; NREM stage 1 is a transitional phase between full wakefulness and sleep, and is characterized by mixed-frequency waves; sleep spindles (12–14 Hz) typify NREM stage 2; and delta waves (1–4 Hz) distinguish NREM stages 3 and 4, which are known collectively as slow-wave sleep. Sleep spindles and delta waves rise and fall in concert with a yet slower (<1 Hz) oscillatory pattern known as cortical slow oscillations — it is these patterns that Marshall *et al.*⁶ have studied.

The authors asked 13 volunteers to learn 46 word-pairs in a training session before they went to sleep. Electrodes were then placed on the scalps of the sleeping subjects, and fluctuating electrical potentials were applied to induce cortical slow oscillations. Remarkably, the morning after this treatment, the subjects demonstrated enhanced recall of the word-pairs compared with their performance the morning after receiving a sham stimulation. This

effect was not seen for all forms of memory. For example, no improvement was observed for a motor-skill procedural memory task — learning to trace around a shape by looking at its image in a mirror. Furthermore, the memory boost occurred only if the electrical stimulation matched the oscillation frequency of cortical slow waves (0.75 Hz) and not when the stimulation was at the theta frequency (5 Hz) of REM sleep. The time chosen for the stimulation was also crucial — the authors observed no effect on recall if stimulation was during the last 45 minutes of the night instead of the first 45 minutes.

The actual increase in words recalled was small; averaged over the whole group, the subjects remembered only 1.8 more words after stimulation than they did after the sham treatment. But the improvement in subjects’ recall compared with their performance during the training session was statistically significant — on average, after stimulation, 4.8 more words were remembered by the group in the morning than had been recalled the previous night, compared with just 2.1 more words in the control experiments. These increases probably do not reflect a genuine overnight improvement, because they were based on comparisons with performance during the training session (when the words would have been imperfectly memorized), not after it (when peak performance would be expected). Nevertheless, sleep-dependent consolidation, enhanced by direct electrical stimulation, does seem to stabilize memories. But why does this happen?

The authors began electrical stimulation of the subjects’ brains 4 minutes after the volunteers had entered stage 2 of NREM sleep — that is, 5 to 10 minutes before they would normally be expected to progress into slow-wave sleep; the stimulation was applied in five 5-minute periods, separated by stimulation-free pauses lasting 1 minute each. Marshall and colleagues’ analysis of the stimulation-free interludes revealed greater electrical activity in the subjects’ brains at slow-oscillation frequencies (up by 60%) and at sleep-spindle frequencies (up by 51%), as well as a 37% increase in the amount of time spent in slow-wave sleep, compared with sham stimulation in the same subjects. Arguably, any of these effects could underlie the observed enhancement of memory consolidation. Indeed, sleep spindles⁷ and the duration of slow-wave sleep⁸ have previously been implicated in memory consolidation. But

Marshall *et al.*⁶ argue that the spindles are the key players. Spindles produce large influxes of calcium ions into cortical neurons, where they can trigger molecular cascades known to strengthen the connections between neurons — which presumably leads to memory consolidation⁹.

Might the effects⁶ of electrical stimulation be caused by alterations in the levels of neuromodulatory chemicals? The authors argue that they are not. They measured blood levels of some of these compounds — cortisol, growth hormone and noradrenaline — and reported no changes from normal as a result of the stimulation. Other neuromodulators were not studied, most notably acetylcholine, which is known to be involved in memory formation. This crucial point requires further exploration, as earlier reports from the same group indicate that low levels of acetylcholine⁵ and cortisol¹⁰ are required for sleep-dependent consolidation of memories.

Such uncertainty indicates a need for continued research into sleep-dependent memory consolidation. But rapid resolution will not come soon, as several complicating issues remain unresolved. First, many different memory systems exist — those involved in memorizing word-pairs may be different from those required for learning to ride a bicycle, or those concerned with recalling the details of an emotional event. Memory consolidation in different systems clearly correlates with different stages of sleep, and possibly even with different components of those stages. Second, consolidation seems to be a series of events, but we do not know which of these are sleep-dependent (within any given memory system). Finally, our understanding of the cellular and molecular mechanisms underlying waking memory consolidation is still highly primitive. It is time for sleep researchers to tackle these basic issues. ■

Robert Stickgold is at the Center for Sleep and Cognition, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02215, USA.
e-mail: rstickgold@hms.harvard.edu

1. Peigneux, P. *et al.* *Neuron* **44**, 535–545 (2004).
2. Walker, M. P., Stickgold, R., Jolesz, F. A. & Yoo, S. S. *Cereb. Cortex* **15**, 1666–1675 (2005).
3. Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F. & Thompson-Schill, S. L. *Curr. Biol.* **16**, 1290–1294 (2006).
4. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. *Nature* **430**, 78–81 (2004).
5. Gais, S. & Born, J. *Proc. Natl Acad. Sci. USA* **101**, 2140–2144 (2004).

*This article and the paper concerned⁶ were published online on 5 November 2006.

6. Marshall, L., Helgadottir, H., Mölle, M. & Born, J. *Nature* **444**, 610–613 (2006).
7. Fogel, S. M. & Smith, C. T. *J. Sleep Res.* **15**, 250–255 (2006).
8. Stickgold, R., Whidbee, D., Schirmer, B., Patel, V. & Hobson, J. A. *J. Cogn. Neurosci.* **12**, 246–254 (2000).
9. Sejnowski, T. J. & Destexhe, A. *Brain Res.* **886**, 208–223 (2000).
10. Plihal, W. & Born, J. *NeuroReport* **10**, 2741–2747 (1999).

DEVICE PHYSICS

A terahertz modulator

Daniel Mittleman

Tiny metal resonators can be used to create a material with tunable responses to an applied voltage. Combined with a semiconductor substrate, they can be used to control technologically promising terahertz radiation.

Electromagnetic radiation with frequencies lying between the microwave region and the infrared — so-called terahertz radiation — holds great promise for imaging and sensing applications. It is non-ionizing, and therefore causes less damage to biological tissue than conventional, higher-energy X-rays. It penetrates plastics and clothing, but not metal, and so is ideal for security screening and non-contact testing or inspection. But to realize the full potential of terahertz radiation, more sophisticated techniques for its generation, manipulation and detection are required. In this issue, Chen *et al.* (page 597)¹ fill an important gap in our capabilities. They report the development of an efficient, electrically driven modulator for terahertz signals that functions at room temperature.

This work builds on exciting developments in the field of metamaterials. These are materials engineered to have electromagnetic responses that are impossible in naturally occurring materials, such as a negative refractive index. The refractive index of a material, n , is a measure of the speed of light in the material, and is given by $n = (\mu\epsilon)^{1/2}$, where μ is the material's 'permeability' to magnetic fields, and ϵ its 'permittivity' to electric fields. All naturally occurring materials have positive μ ; transparent materials have positive ϵ , too. In these normal materials, therefore, the refractive index is a real and positive number.

In 1968, the Soviet physicist Victor Veselago showed² that a hypothetical material with negative ϵ and μ would also have a real refractive index, meaning that light waves could propagate through the material, but would behave as though its refractive index were negative. This material would have unusual and potentially valuable properties. A flat slab of negative-index material, for example, would focus light in much the same way as a curved slab of ordinary material (a lens), but with a smaller focal spot.

Although negative-index materials do not violate any laws of physics, the absence of a medium with negative μ confined the idea to the realm of speculation. But in the late 1990s, John Pendry found that, by assembling a collection

of appropriately designed metallic structures, a material can be fabricated that has both negative ϵ and negative μ for incident electromagnetic radiation of a particular frequency^{3,4}. Furthermore, if the metal structures are each much smaller than the wavelength of the incident radiation, the radiation interacts with them not individually, but collectively, according to their average properties. These are the engineered materials now known as metamaterials.

Engineering a material with negative ϵ was easy: this equates to opacity, a property of all metals for incident radiation below a certain frequency. It was necessary to show only that a discrete set of thin metallic structures could mimic this property of the bulk metal³. The more difficult task was achieving a negative μ . It turned out that this could be done using a pair of concentric metallic rings with gaps that prevent current from circulating. Because these rings are both capacitors (they store electric charge) and inductors (they induce magnetic fields that self-sustain any current flowing through them), the presence of gaps leads to a resonant response, with charge accumulating alternately on one side of the gap and then the other, sloshing back and forth through the rings rather as a mass vibrates back and forth on a spring. At frequencies near the characteristic frequency of this resonant electron flow, ϵ and μ can vary dramatically as a function of frequency. Indeed, either one can become negative if the resonance is strong enough⁴.

The development of the split-ring-resonator concept was significant not only because it permits a negative refractive index, but more generally because it represents a new technique for 'designing' the optical response of a medium. The first experimental demonstration of a negative index⁵, along with nearly all research into metamaterials until now, was performed in the microwave regime. This region encompasses gigahertz frequencies below the terahertz regime, with wavelengths of several millimetres or longer. It was almost immediately recognized, however, that the approach could be extended into the shorter-wavelength, terahertz regime simply by shrinking

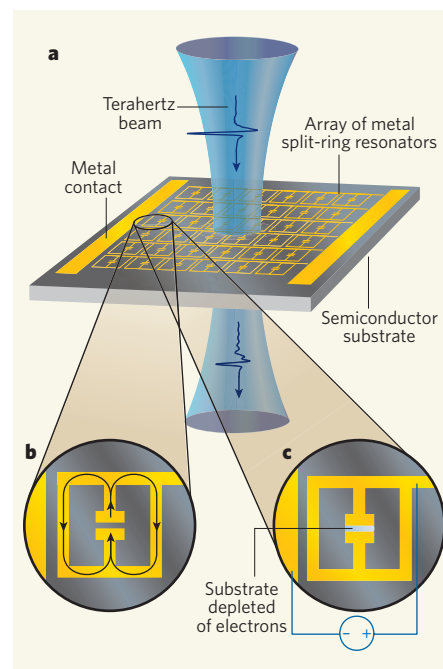


Figure 1 Chen and colleagues' resonant metamaterial modulator¹. **a**, The tunable resonant response of an array of metallic split-ring resonators on a thin semiconductor substrate can be exploited as a modulator, letting more or less of an incident radiation beam through. **b**, When no external voltage is applied, the doped semiconducting substrate conducts. Current circulates around the two lobes of the resonator in response to the applied terahertz field. **c**, When an external voltage is applied, the semiconductor ceases to conduct. Current can no longer flow through the ring gap; this gap instead behaves as a capacitor, storing up charge. The capacitance gives rise to a resonant behaviour at a frequency in the terahertz range, in which the electrons slosh back and forth between the upper and lower gap electrodes. The incident terahertz radiation drives this resonance efficiently, leading to the transfer of more energy from the beam to the electrons of the substrate. Thus, less radiation is transmitted through the device.

the size of the individual metallic components so that they remained smaller than the incident radiation's wavelength. At 1 THz, this is 300 μm , so the fabrication of a negative-index medium requires the technically challenging construction of three-dimensional objects with micrometre-scale features. Two-dimensional patterns on that scale, on the other hand, can be easily generated using conventional photolithography, and these patterns can be designed to exhibit a strong resonance in either ϵ or μ at any frequency of interest.

Chen *et al.*¹ use the split-ring-resonator concept as the basis for a metamaterial that provides a resonant response in ϵ — although not a negative refractive index — in the terahertz range. Furthermore, they have shown that this resonance can be externally controlled, and therefore can be exploited as a modulator for controlling the transmission of terahertz electromagnetic radiation.

The principle of this device's operation is simple and elegant (Fig. 1). An array of sub-wavelength metallic ring resonators is deposited on top of a thin, lightly doped semiconductor layer and illuminated with a beam of terahertz radiation. The electrons in the semiconductor substrate effectively short-circuit the gap in the split rings, driving the capacitance of the rings to zero and damping their normal resonant response. But when a negative voltage is applied to the metal structures, the electrons in the substrate beneath are repelled away from the gap. As electrons become depleted in the gap, current can no longer flow effectively and the gap behaves as a capacitor, storing electric charge. In this case, the resonance of the ring structure re-emerges, which gives rise to a pronounced change in the optical properties of the array at frequencies near its resonant frequency.

The authors' results¹ are impressive. At the design frequency, the on-off transmission ratio of this device is about 0.5 — more than ten times better than state-of-the-art, electrically operated modulators in this frequency range^{6,7}.

The new design is simpler to fabricate, and the modulator works at room temperature. This excellent performance comes from the exquisite sensitivity of a metamaterial's response to the precise properties of its resonant substructures. The ability to electrically switch the properties of a metamaterial by fabricating it on a semiconductor substrate provides a new method for active control of terahertz devices.

Naturally, challenges remain. Most obviously, larger on-off transmission ratios will be required for many applications. The modulation speed, which is only a few kilohertz, must be increased significantly. The authors point out that both of these difficulties can be addressed by optimizing both the pattern of the metal structures and the properties of the substrate. The properties of the device are also dependent on the direction of the electric-field vector (the polarization) of the incident radiation: the resonant behaviour of the split rings relies on this electric field driving current across the gaps, all of which are oriented in one particular direction. Designs of future

devices may seek to eliminate or alternatively exploit this polarization sensitivity. Finally, one can imagine a structure for which not only the transmitted intensity, but also the resonant frequency, is externally controllable.

Whatever the next stage of development might be, the resonator structures described by Chen *et al.*¹ open a new and promising set of possibilities for the active control of terahertz radiation, with all the potential that it holds. ■

Daniel Mittleman is in the Department of Electrical and Computer Engineering, Rice University, 6100 Main Street, Houston, Texas 77005, USA.
e-mail: daniel@rice.edu

1. Chen, H.-T. *et al.* *Nature* **444**, 597–600 (2006).
2. Veselago, V. G. *Sov. Phys. Uspekhi* **10**, 509–514 (1968).
3. Pendry, J. B., Holden, A. J., Stewart, W. J. & Youngs, I. *Phys. Rev. Lett.* **76**, 4773–4776 (1996).
4. Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. *IEEE Trans. Microwave Theory Tech.* **47**, 2075–2084 (1999).
5. Shelby, R. *et al.* *Science* **292**, 77–79 (2001).
6. Kleine-Ostmann, T., Dawson, P., Pierz, K., Hein, G. & Koch, M. *Appl. Phys. Lett.* **84**, 3555–3557 (2004).
7. Kersting, R., Strasser, G. & Unterrainer, K. *Elec. Lett.* **36**, 1156–1158 (2000).

SYSTEMS BIOLOGY

Many things from one

John R. S. Newman and Jonathan S. Weissman

Cells of the same type can generate diverse sets of physiological traits from a single set of genes. Part of this diversity could stem from 'noise' that arises from variations in the way proteins are expressed.

The census bureau will tell you that the typical American family has 2.1 children, but there are no families (we hope) that precisely match this mean. Similarly, biologists have come to realize that population-based measurements can obscure critical information about cell-to-cell diversity. The use of fluorescent proteins as tags to look at the abundances of proteins in single cells has revealed that, at least for microorganisms, individual protein levels can vary considerably, even for genetically identical populations grown under uniform conditions.

On page 643 of this issue, Sigal and colleagues¹ present a view of what protein variation may look like in human cells*. They reveal that this variation (or 'noise') can persist over several cell generations. Additionally, for proteins belonging to the same biochemical pathway, variation seems to be larger between cells than within cells. To the extent that protein abundance and activity are correlated, this variation might contribute to the ability of cells to generate a diversity of behaviours from a single set of genes (genotype).

Variation has attracted considerable interest,

particularly because it can affect cells differentially. It can be an obstacle to the precise functioning of cells, by causing levels of biological molecules to deviate from their optima, and can degrade biological signals. But noise can also be exploited by allowing cells to switch between expression states, or, more speculatively, by generating diversity in physiological characteristics (phenotypes).

Studies using microorganisms have identified at least two types of noise. Extrinsic noise results from intercellular variations in the levels of components of the pathways that regulate gene expression. Intrinsic noise, however, arises from the random production and/or destruction of messenger RNAs and proteins that is due to chance interactions occurring in cells. These interactions would persist even if every cell were otherwise identical. Extrinsic noise tends to dominate for proteins present in high amounts in a cell, whereas intrinsic noise dominates when proteins, and their corresponding mRNAs, are present in low copy number. In contrast to our understanding of variation in microorganisms, much less is known about the origins and consequences of variation in the cells of multicellular organisms. A priori one might speculate that, in

these cells, the cost of noise might be greater because of the interdependence of cells within tissues or organs. This, together with the fact that there are higher numbers of mRNA molecules per cell, which tends to average out the randomness of the production and destruction of individual messages, suggests that total noise might be lower than that observed in microorganisms.

To explore this issue directly, Sigal and colleagues¹ created a series of strains of a human cancer cell line in which DNA encoding yellow fluorescent protein (YFP) was inserted into the genome in the middle of genes encoding various proteins (one insertion per strain). The genes therefore maintained their natural regulation, but encoded fluorescent fusion proteins. As the cells grew, time-lapse microscopy was used to monitor their overall fluorescence (and thus the abundance of the respective protein). The images were 'synchronized' with respect to cell divisions, eliminating a major source of variation that was not pertinent to these studies. The authors found that the coefficient of variation (standard deviation/mean) is 10–30%, similar to measurements in microorganisms².

To explore the level of intrinsic and extrinsic noise, the authors succeeded in obtaining a strain in which both copies of the gene that encodes the RPL5 protein were tagged, one with the DNA encoding YFP and one with DNA encoding a red fluorescent protein (RPL5 is found in a multi-subunit complex called the ribosome, which is involved in protein synthesis). Such a strain is useful because it makes it possible to distinguish between extrinsic and intrinsic noise (observed as intercellular and intracellular differences in fluorescence,

*This article and the paper concerned¹ were published online on 19 November 2006.

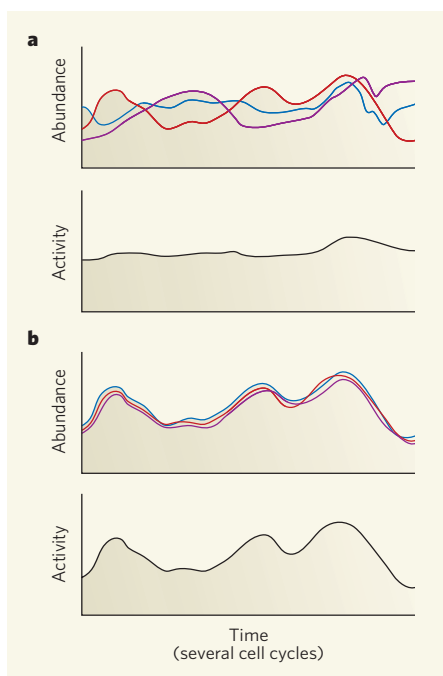


Figure 1 | Potential link between coherent protein variation and phenotype. **a**, In a three-component pathway, when the variation in the levels of each component is uncorrelated (top), the total pathway activity (bottom) is unrelated to the variation shown by any single component. This would mean that the phenotypic outcome of this pathway would be similar among a population of cells. **b**, When variation is correlated (top), the total pathway activity (bottom) varies in direct proportion to the variation shown by any single component. Sigal and colleagues¹ show that the abundances of ribosomal proteins tend to vary in concert. Extension of this finding to other proteins, particularly those for which high (or low) levels are maintained over several cell cycles, suggests that extrinsic noise may generate phenotypic diversity across a population of cells.

respectively). In double-tagged cells, the levels of yellow and red fluorescence tended to fluctuate in concert, showing that extrinsic noise is a major component of total noise. This is in line with what one would expect given the greater numbers of mRNA molecules in human cells and the study's bias towards abundant proteins.

An extension of this double-tagging approach shows that, at least for ribosomal subunits, proteins that are involved in similar processes show similar variation, whereas those participating in different biological pathways do not always do so. Additionally, Sigal *et al.* looked at the length of time proteins spend above or below a population mean. For the 20 fusion proteins used in their study, they find that proteins require 0.8 to 2.5 cell cycles to 'relax' towards the average.

The picture that emerges from their observations is one in which the coherent induction or repression of a pathway occurs such that high or low protein levels can be maintained for several days. This phenomenon, particularly if it proves to be general, is provocative because it

hints that proteins that must act together will be induced or repressed in a coordinated manner in individual cells, possibly even generating phenotypic diversity among cells (Fig. 1). So how might this long-term action occur, and what are the probable physiological consequences? One proposed model is that random fluctuations in regulatory components are amplified as a signal cascades down a pathway. Alternatively, long-lived responses may result from self-reinforcing mechanisms, such as positive feedback, that lead to slow rates of conversion between distinct activation states of gene networks.

Harder to gauge is the phenotypic impact of this extrinsic variation. A possible benefit of coherent variation could be to ensure that processes that are energetically expensive, such as protein secretion, are restricted to a few cells³, rather than having many cells incur a high fixed cost to produce a set of proteins that are underused. Also, by analogy to microorganisms, where variation is postulated to generate phenotypic heterogeneity⁴, one might speculate that noise could allow a subpopulation of cancer cells to evade endogenous or exogenous attempts to inhibit unrestrained cell growth.

At least three issues await further investigation. First, to what extent do the tissue-culture conditions reflect what occurs *in vivo*? Although tissue culture is an artificial environment that might generate noise that doesn't

occur *in vivo*, cells typically divide much more slowly *in vivo*, which could cause deviations to persist for even longer timescales than Sigal *et al.* report. Second, to what extent will the observations reported here prove general? As the authors point out, the set of proteins they examine are biased (for example, by their abundance and accessibility for YFP integration). Third, do variations in protein concentrations lead to proportional changes in their activities? After all, cells have many mechanisms for controlling activities. These include modifying a protein after it has been made, and regulating the protein's cellular localization and the availability of key partners, all of which might act to suppress variation. More broadly, can we link heterogeneity in protein concentrations or activities to phenotypic differences? Although these questions show that much remains to be done, it is clear that these and related topics can now be explored in the cells of multicellular organisms. ■

John R. S. Newman and Jonathan S. Weissman are in the Department of Cellular and Molecular Pharmacology, and at the Howard Hughes Medical Institute, University of California, San Francisco, Byers Hall 403B, 1700 4th Street, San Francisco, California 94158-2330, USA. e-mail: jrsnewman@alum.mit.edu

1. Sigal, A. *et al.* *Nature* **444**, 643–646 (2006).
2. Newman, J. R. S. *et al.* *Nature* **441**, 841–846 (2006).
3. Michaelson, J. *Lab. Invest.* **69**, 136–150 (1993).
4. Balaban, N. Q. *et al.* *Science* **350**, 1622–1625 (2004).

HYDROLOGY

Water from on high

Dennis P. Lettenmaier and James S. Famiglietti

Data on changes in water storage in the Congo basin show how GRACE, a pair of satellites designed to record variations in Earth's gravitational field, is benefiting the study of the planet's water cycle.

Hydrologists yearn above all to close Earth's water balance. This is essentially a housekeeping task: to measure precisely where and in what quantities Earth stores water, and how the water moves between those stores. Constant changes, both natural and anthropogenic, to the global water cycle make this calling especially exacting, and particularly important¹. Writing in *Geophysical Research Letters*, Crowley *et al.*² further the cause with measurements of land water storage in the Congo river basin in central Africa. Their observations provide further evidence of a 'see-saw' anti-correlation with similar measurements in the Amazon basin of South America. Most significant of all, however, is how these data were acquired: through sensitive space-based measurements of Earth's gravitational field, courtesy of the Gravity Recovery and Climate Experiment, or GRACE for short.

On Earth's land surfaces, the main moisture

fluxes are precipitation, evaporation and transpiration (evaporation from plants, often lumped together with evaporation as evapotranspiration) and run-off. The balance of these fluxes is reflected in changes in moisture storage in both surface and subsurface stores; not only lakes, reservoirs, wetlands and rivers, but also snow and glaciers, and groundwater and soil moisture. Measuring each of the flux and storage terms presents its own challenges. But evapotranspiration — for which little more than model estimates exist, despite various efforts to tie these to local observations³ — and changes in storage are the most problematic of all.

A strategy that has been used to sidestep these problem terms and estimate the terrestrial water budget over large areas is to perform averages over many years. In this case, the effect on the overall balance of interannual variations in terrestrial storage is small; but on

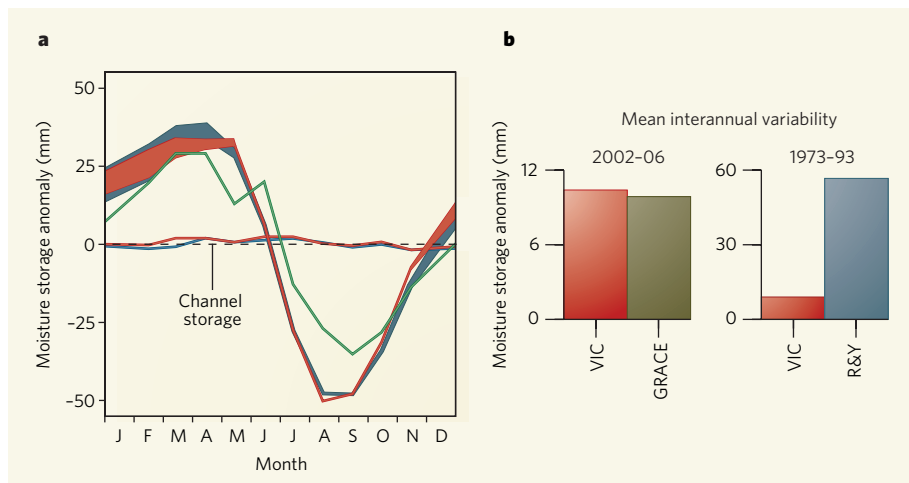


Figure 1 | Estimates of terrestrial water storage in the Mississippi river basin. **a**, The month-by-month deviation of moisture storage from the annual mean as calculated from the variable infiltration capacity (VIC) model¹⁰ using data for 1950–2005 (blue) and just 2002–06 (red), for comparison with GRACE data from the same period (green). Shading indicates the contribution of seasonal snowpack; the much smaller contribution of storage in channels (rivers, lakes, and so on) is shown in separate lines. The 1950–2005 and 2002–06 model estimates are closely coincident, indicating that the GRACE period is typical of the longer record. **b**, The VIC model and GRACE data agree closely on the variability of terrestrial moisture storage between years; an earlier study (R&Y)¹¹ based on the atmospheric moisture budget put this variability much higher. (Figure courtesy of J. Adam, Univ. Washington; GRACE data processing by T. H. Syed, Univ. California, Irvine.)

the downside, one acquires no information about the dynamics of the contributing terms. Alternatively, one can construct a total atmospheric water budget. Here, water loss through precipitation, change in moisture storage and atmospheric convergence — the net lateral flow into the atmospheric column over which the budget calculations are performed — are all known quantities. Evapotranspiration is the missing ‘closure term’ in the atmospheric budget, and this term can be plugged back into the terrestrial water balance. This approach provides a quasi-independent estimate of evapotranspiration, but can be applied only to fairly large river basins, over which errors in estimates of atmospheric convergence are small.

The GRACE satellites were launched by NASA and the German Aerospace Center, DLR, in March 2002, and provide a new source of information to assist in unravelling the complexities of the terrestrial water balance. The two satellites measure month-to-month changes in surface and subsurface water storage through the effect of these parameters on Earth’s gravitational field. Although GRACE’s spatial resolution (nominally around 300,000 km²) is coarse, it provides observations that are independent of atmospheric and surface conditions. Errors are thus globally consistent.

Crowley *et al.*² use GRACE data to assess variations in terrestrial moisture storage in the Congo basin of central Africa between 2002 and 2006. Their results show that interannual variations were out of phase with those in the Amazon basin during this period, confirming an anti-correlation inferred earlier from precipitation and streamflow anomalies⁴. This see-saw behaviour is thought to be related to the so-called Gill mechanism, which holds

that, near the Equator, persistent atmospheric uplift (and associated vapour condensation, and so precipitation) in one area will be compensated by atmospheric subsidence, and so precipitation deficits, in another⁵.

Crowley and colleagues’ data also show a downward trend in Congo terrestrial storage over the short period surveyed. This might well be just a manifestation of decadal-scale variations in precipitation already indicated by streamflow records⁴, but also indicates the relatively high magnitude of interannual variations in terrestrial storage when compared with seasonal variations.

Other researchers have also started to use the GRACE record to estimate seasonal and interannual variations in the terrestrial water cycle. The data have been applied to evaluating whether global climate models can capture the seasonal cycle of continental moisture storage⁶. Comparison with observed soil-moisture and groundwater levels in Illinois has shown that the GRACE record tracks moisture levels, and that the interseasonal and interannual variations in it are due roughly equally to soil moisture and shallow groundwater⁷. The atmospheric and land-surface water budgets from GRACE have been combined to estimate the run-off from large river basins, assuming that this is the residual term once atmospheric moisture convergence and terrestrial and atmospheric storage changes have been taken into account⁸. And seasonal variations in moisture storage in 53 river basins covering most of the global land area have been estimated, along with sources of error and their variations with basin area, latitude and shape⁹.

The Mississippi river basin provides an interesting case study for the use of the GRACE

data. Soil moisture and groundwater — rather than lakes and other forms of surface storage — dominate terrestrial water storage in the Mississippi basin. But in this region, high-quality climatological data offer an opportunity to estimate the terms in the terrestrial water budget using land-surface models¹⁰. A comparison of seasonal and interannual variations from model calculations and GRACE data throws up some interesting disparities (Fig. 1). First, model estimates of seasonal variability in soil moisture and snow storage are greater than those from the GRACE data for the same period. This might indicate that the groundwater contribution is modest over the Mississippi basin, as land-surface models do not represent groundwater explicitly. But this remains conjecture, as other researchers have suggested that groundwater may contribute substantially to interannual moisture-storage variability, over at least part of the Mississippi basin⁷. Second, the mean interannual variability inferred from the atmospheric water budget¹¹ is several times that obtained from either the GRACE data or model soil-moisture results, possibly through accumulated error in the estimated atmospheric convergence¹².

At this point, the GRACE period of record is too short to answer questions about the relative contributions of soil-moisture and groundwater (or perhaps other) effects to terrestrial-storage variations. The data do, however, provide a viable means of understanding the sources of these variations, with their various scientific and practical implications. Their role in pinpointing as-yet poorly understood curiosities such as Crowley and colleagues’ Amazon–Congo anti-correlation² could also prove invaluable. Our insight into such phenomena will no doubt improve as the GRACE record is extended.

Dennis P. Lettenmaier is in the Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, USA. James S. Famiglietti is in the Department of Earth System Science, University of California, Irvine, Irvine, California 92697, USA. e-mails: dennisl@u.washington.edu; jfamigli@uci.edu

1. Water Study Group, US Global Change Research Program www.usgcrp.gov/usgcrp/Library/watercycle/wcsgreport2001/default.htm (2001).
2. Crowley, J. W. *et al. Geophys. Res. Lett.* **33**, doi:10.1029/2006GL027070 (2006).
3. Betts, A. K. *et al. J. Clim.* **12**, 1601–1618 (1999).
4. Eltahir, E. A. B. *et al. Geophys. Res. Lett.* **31**, doi:10.1029/2004GL021160 (2004).
5. Gill, A. E. Q. *J. R. Meteorol. Soc.* **106**, 447–462 (1980).
6. Swenson, S. C. & Milly, P. C. D. *Wat. Resour. Res.* **42**, doi:10.1029/2005WR004628 (2006).
7. Swenson, S. C. *et al. Geophys. Res. Lett.* **33**, doi:10.1029/2006GL026962 (2006).
8. Syed, T. H. *et al. Geophys. Res. Lett.* **32**, doi:10.1029/2005GL024851 (2005).
9. Seo, K.-W. *et al. Wat. Resour. Res.* **42**, doi:10.1029/2005WR004255 (2006).
10. Maurer, E. P. *et al. J. Clim.* **15**, 3237–3251 (2002).
11. Ropelewski, C. F. & Yarosh, E. S. *J. Clim.* **11**, 2180–2190 (1998).
12. Trenberth, K. E. & Guillemot, C. *J. Clim. Dynam.* **14**, 213–231 (1998).

BRIEF COMMUNICATIONS

Wood used by Stradivari and Guarneri

The material used by the old masters to make exquisite violins may have been chemically manipulated.

Whether or not the great Italian violin-makers used wood that had been chemically processed in order to preserve it and enhance the instrument's sound quality has long been a contentious issue^{1,2}. Here we use nuclear magnetic resonance and infrared spectroscopy to analyse organic matter in wood taken from antique instruments made by Stradivari and Guarneri del Gesù. Our results indicate that the wood used by the masters could indeed have been chemically treated, a technique that may inspire an approach to violin making that is more chemistry-based.

Maple wood samples from the interior of the instruments' back plates were obtained as thin shavings during the repair of cracks. The instruments were a violin by Stradivari, dated 1717; a cello by Stradivari (1731); a violin by Guarneri del Gesù (1741); a violin by Gand-Bernardel Paris (1840s); and a viola by Henry Jay of London (1769). For comparison, we used recent tone woods from Bosnia and central Europe.

Figure 1a shows ¹³C solid-state nuclear magnetic resonance (NMR) spectra of the different wood samples, recorded (see supplementary information) and interpreted as described^{3,4}. The spectra differ at the following points: the small peaks attributed to the acetyl carbons at 18 and 170 p.p.m.; the small methoxy peak at 56 p.p.m.; and the main lignin peaks at 135 and 155 p.p.m. The most significant difference is shown by the Guarneri sample as a decrease in all of these peaks. The spectrum of

the Stradivari violin shows close similarities to that of the Guarneri, whereas the deviations of the Stradivari cello wood from the recent Bosnian maple are only slightly different from those of the old English and French instruments, and from those generated by boiling and baking the Bosnian maple sample (for details, see supplementary information).

Attenuated total reflection Fourier-transform infrared (FTIR) spectra⁵ were obtained for recent Bosnian maple, the Guarneri sample and the Stradivari violin sample (Fig. 1b). The differences between these spectra are obvious at several peaks, particularly in the carbonyl region at 1,730 cm⁻¹ to 1,650 cm⁻¹ and at 1,237 cm⁻¹. The increase in absorption at around 1,650 cm⁻¹ may be caused by the formation of quinones from lignin by oxidation⁶. The spectra of both violins also differ from that of the Stradivari cello, which in turn differs only slightly from those of Jay and Gand-Bernardel and from the boiled-baked Bosnian maple. (See supplementary information for spectra and statistical analysis of their significance.) We found no evidence of potassium



What gives this Stradivari cello its characteristic sound?

silicate⁷ in any of our antique samples.

By these methods, the violin by Guarneri and, to a slightly lesser degree, the Stradivari stand out against the other standards. Our results support the idea that chemical treatments, such as oxidation and hydrolysis, were used in making these violins and, to a much lesser extent, the Stradivari cello. In the case of the cello, the English viola and the French violin, natural ageing cannot be ruled out as a cause of the chemical changes. The effects may also be related to the unusual mineral composition of the wood in these instruments, which remains to be investigated.

However, the observed differences are likely to have originated from a regional practice of wood preservation that affected the mechanical and acoustical properties of the wood^{2,8}. Our case studies may inspire a more chemistry-based approach to violin-making.

Joseph Nagyvary*, Joseph A. DiVerdi†, Noel L. Owen‡, H. Dennis Tolley§

*Department of Biochemistry, Texas A&M University, College Station, Texas 77843, USA
e-mail: nagyvary@nagyvaryviolins.com

†Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, USA
Departments of ‡Chemistry and Biochemistry, §Statistics, Brigham Young University, Provo, Utah 84602, USA

1. Gough, C. *Phys. World* <http://physicsweb.org/articles/world/13/4/8> (April 2000).
2. Nagyvary, J. *Edu. Chem.* **42**, 96–98 (2005).
3. Kolodziejewski, W., Frye, J. S. & Maciel, G. E. *Anal. Chem.* **54**, 1419–1424 (1982).
4. Kosikova, B., Hricovini, M. & Cosentino, C. *Wood Sci. Technol.* **33**, 373–380 (1999).
5. Orton, C. R., Parkinson, D. Y., Evans, P. D. & Owen, N. L. *Appl. Spectrosc.* **58**, 1265–1271 (2004).
6. Anderson, E. L., Pawlak, Z., Owen, N. L. & Feist, W. C. *Appl. Spectrosc.* **45**, 641–647 (1991).
7. Sacconi, S. F. I. "Segreti" di Stradivari (Libreria del Convegno, Cremona, 1972).
8. Haines, D. W. *Catgut Acoust. Soc. Newslett.* **33**, 19–23 (1980).

Supplementary information accompanies this communication on Nature's website.

Received 18 July; accepted 8 November 2006.

Competing financial interests: declared none.
doi:10.1038/444565a

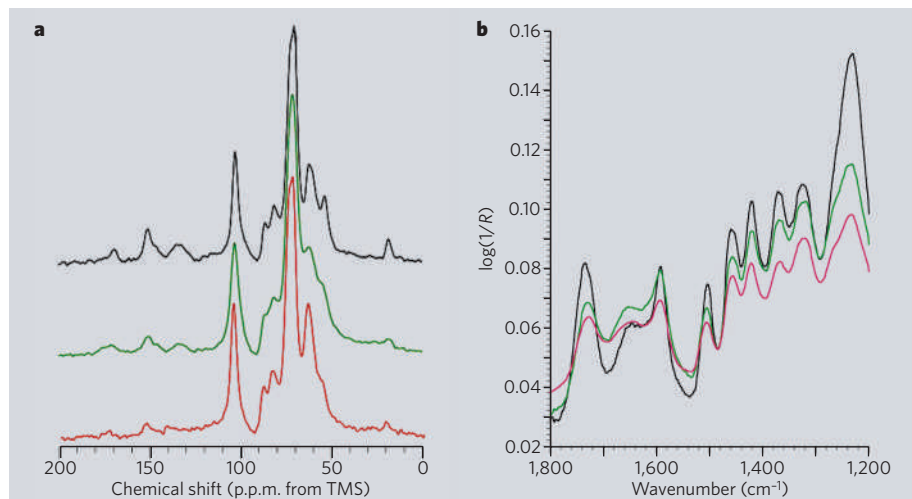


Figure 1 | Spectroscopy of maple wood samples from antique violins. **a**, ¹³C solid-state NMR spectra. TMS, tetramethylsilane standard. **b**, Attenuated total reflection Fourier-transform infrared spectra. Black trace, recent untreated Bosnian maple; green, violin of Stradivari; red, violin of Guarneri del Gesù. Conditions and statistical analysis are described in supplementary information.

OENOLOGY

Red wine procyanidins and vascular health

Regular, moderate consumption of red wine is linked to a reduced risk of coronary heart disease and to lower overall mortality¹, but the relative contribution of wine's alcohol and polyphenol components to these effects is unclear². Here we identify procyanidins as the principal vasoactive polyphenols in red wine and show that they are present at higher concentrations in wines from areas of southwestern France and Sardinia, where traditional production methods ensure that these compounds are efficiently extracted during vinification. These regions also happen to be associated with increased longevity in the population.

High consumption of polyphenols inhibits atherosclerosis in experimental models^{3,4}. Red-wine polyphenols induce endothelium-dependent dilatation of blood vessels and suppress the synthesis of endothelin-1 (ET-1), a peptide that has a vasoconstricting effect^{5–7}, and this may account for their anti-atherosclerotic activity. However, there is a lack of consensus on the protective effects of red wine, which may be due to variability in vasoactive constituents in different wines.

Red-wine polyphenols are a complex mixture of flavonoids (mainly anthocyanins and flavan-3-ols) and non-flavonoids (such as resveratrol and gallic acid). Flavan-3-ols are the most abundant, with oligomeric and polymeric procyani-

dins (condensed tannins) often representing 25–50% of the total phenolic constituents⁸.

We used cultured endothelial cells to identify the most potent vasoactive polyphenols in red wine (for methods, see supplementary information). These were shown by high-performance liquid chromatography with mass spectrometry² to be straight-chain B-type oligomeric procyanidins (OPCs) (tetra-epicatechin gallate, $m/z = 1,305$; procyanidin trimer-gallate, $m/z = 1,017$; procyanidin tetramer, $m/z = 1,153$; and pentamer-gallate, $m/z = 1,593$; see supplementary information).

Total polyphenols and OPC content of each wine correlated with the suppression of ET-1 synthesis (Fig. 1a, b). However, the linear regression plot for total polyphenols intercepted the y-axis at about 5 mM, which is consistent with most polyphenols (anthocyanins, catechins and resveratrol) lacking vasoactivity at the concentrations found in wine^{5,7}.

To investigate how the OPC content of red wines from a particular region might relate to mortality in that region, we compared wines produced in areas of increased longevity (as an index of overall good health) with a broad selection of wines from different countries. People living in Nuoro province, Sardinia, have high longevity, particularly men⁹. In France, there are marked regional variations in mortal-

ity from coronary heart disease. We used the 1999 census data to identify unusual patterns of ageing in France (see supplementary information) and found that there are relatively more men aged 75 or over in the *département* of Gers in the Midi-Pyrenees in southwest France.

Wines from Nuoro and the Gers area have 2–4-fold more biological activity and OPC content than other wines (Fig. 1c, d). This difference remains ($P < 0.001$) when OPC measurements are extended to a wider selection of wines from the Gers area (2.9 ± 0.1 mM, $n = 58$), from France (1.8 ± 0.1 mM, $n = 61$) and from other parts of the world (1.5 ± 0.04 mM, $n = 227$).

Grape seeds are the main source of OPCs but poor solubility, combined with oenological and viticultural factors, influence the amount of OPCs in wine⁸. The higher OPC concentration in wines from southwest France is due to traditional wine-making, which ensures that high amounts of OPCs are extracted, and to the flavonoid-rich grape Tannat, which makes up a large proportion of grapes used to produce local wines in the Gers area but is rarely grown elsewhere.

Absorption of OPCs and their identification in plasma has been demonstrated *in vivo*¹⁰, but little is known about their biological availability and metabolism. Further investigation of OPC-rich wines and foods should provide insight into how vascular function might be optimally maintained.

R. Corder*, **W. Mullen†**, **N. Q. Khan***, **S. C. Marks†**, **E. G. Wood***, **M. J. Carrier***, **A. Crozier†**
*William Harvey Research Institute, Barts and the London, Queen Mary's School of Medicine and Dentistry, London EC1M 6BQ, UK
e-mail: r.corder@qmul.ac.uk

†Plant Products and Human Nutrition Group, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

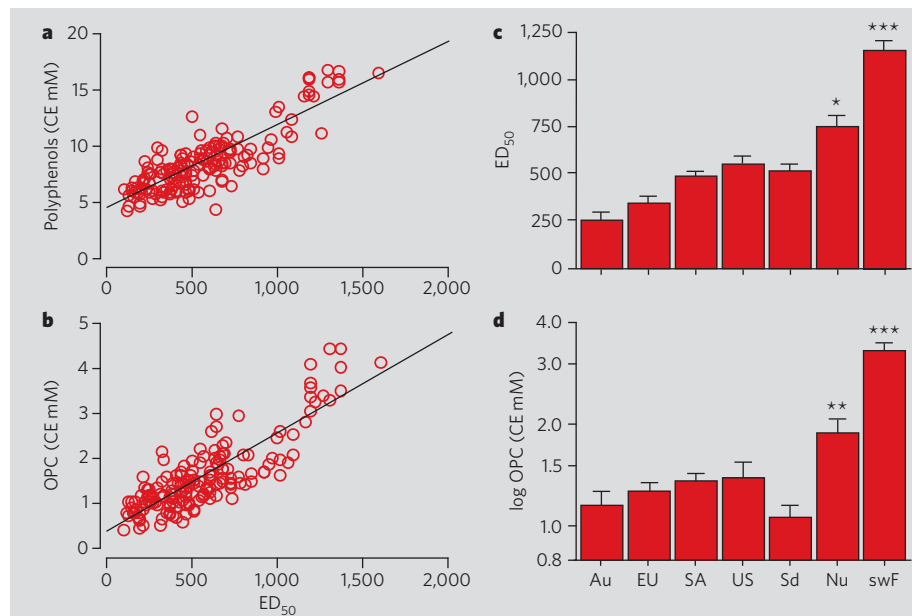


Figure 1 | Relationship between procyanidin content and vasoactive properties of red wine. **a, b,** Total polyphenol (**a**) and oligomeric procyanidin (OPC) (**b**) content correlate with the inhibition of synthesis of endothelin-1, expressed as ED₅₀ (dilution inhibiting by 50%; see supplementary information); $R = 0.84$ for both, $n = 165$. **c, d,** Comparison of inhibition of endothelin-1 synthesis (**c**) with OPC concentration (**d**) of wines from different geographical regions. Au, Australia; EU, France, Greece, Italy or Spain; SA, South America; US, United States; Sd, Sardinia; Nu, Nuoro province, Sardinia; swF, southwest France. CE, catechin equivalents (see supplementary information). *** $P < 0.001$ compared with all the other wines; * $P < 0.01$ compared with the United States, and $P < 0.001$ compared with the other wines; ** $P < 0.02$ compared with the United States and South America, and $P < 0.001$ compared with the other wines.

1. Renaud, S. C., Gueguen, R., Siest, G. & Salamon, R. *Arch. Intern. Med.* **159**, 1865–1870 (1999).
2. Di Castelnuovo, A., Rotondo, S., Iacoviello, L., Donati, M. B. & De Gaetano, G. *Circulation* **105**, 2836–2844 (2002).
3. Kris-Etherton, P. M. & Keen, C. L. *Curr. Opin. Lipidol.* **13**, 41–49 (2002).
4. Auger, C. et al. *J. Agric. Food. Chem.* **52**, 5297–5302 (2004).
5. Fitzpatrick, D. F., Hirschfield, S. L. & Coffey, R. G. *Am. J. Physiol.* **265**, H774–H778 (1993).
6. Hashimoto, M. et al. *Am. J. Cardiol.* **88**, 1457–1460 (2001).
7. Corder, R. et al. *Nature* **414**, 863–864 (2001).
8. Waterhouse, A. L. *Ann. N. Y. Acad. Sci.* **957**, 21–36 (2002).
9. Poulain, M. et al. *Exp. Gerontol.* **39**, 1423–1429 (2004).
10. Shoji, T. et al. *J. Agric. Food. Chem.* **54**, 884–892 (2006).

Supplementary information accompanies this communication on Nature's website.

Received 29 August; accepted 9 November 2006.

Competing financial interests: declared, details accompany the paper on Nature's website.
doi:10.1038/444566a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

ASTROCHEMISTRY

Complex organic matter in Titan's aerosols?

Arising from: G. Israël *et al.* *Nature* **438**, 796–799 (2005)

On 14 January 2005, the Huygens probe entered the atmosphere of Titan after a seven-year interplanetary flight as part of the Cassini mission to Saturn. Huygens carried, among other instruments, an aerosol collection and pyrolysis (ACP) device¹. Its designers, Israël *et al.*², now claim to have detected complex organic matter in two aerosol samples collected at different altitudes (130–35 km and 25–20 km, respectively), on the basis of their detection of ammonia (NH₃) and hydrogen cyanide (HCN) when the sample oven was heated to 600 °C. However, the authors' remarkable conclusions, which would have far-reaching consequences for our understanding of the chemical environment prevailing on Saturn's largest moon, are not supported by their limited data.

The claim by Israël *et al.*² to have detected NH₃ is based on the signal they obtain at a mass-to-charge ratio (*m/z*) of 17 in the 18 mass spectra (see their Fig. 3) recorded for sample 2 after the sample oven had been heated to 600 °C and swept with an isotope of nitrogen (¹⁵N₂). However, with a single exception (at 10:57:27), these signals can be accounted for by methane containing ¹³C (that is, ¹³CH₄), as calculation from the signal at *m/z* 16 (¹²CH₄) and the ¹²C/¹³C ratio (82.3) of Titan's atmospheric methane indicates³; sample 1 apparently produced similar results². Considering the small signals and large error bars indicated in Fig. 3 of ref. 2, it is not valid to conclude that ammonia has been detected in these experiments.

The claim by Israël *et al.*² that HCN could be detected is also unsupported by their data. The signal for *m/z* 27 in the mass spectrum for sample 1 is the same as that from the background (their Fig. 1) and, as the authors point out, the small peak in the spectrum for sample 2 (their Fig. 2) may well be a fragment of ethane or ethylene. The signals above *m/z* 27 of these known constituents of Titan's atmosphere are

obliterated by those from the large amounts of light (*m/z* 28, 29) and heavy nitrogen (*m/z* 30).

These two spectra are dominated by signals at *m/z* 16, 28 and 40, which correspond to the components of the atmosphere (CH₄, ¹⁴N₂ and ⁴⁰Ar). This may have entered the sample cavity through the leaking gaskets, which had caused the loss of all data from the "ambient" and 250 °C experiments² (see section 2 and Fig. 4 of the supplementary information accompanying ref. 2). These leaks would also explain the increase in the signals for sample 2, because it was collected at lower altitude (higher pressure). There is therefore no compelling evidence that any aerosol was collected or that anything was pyrolysed in these experiments.

What the authors mean by the "complex organic matter" they claim to have detected in Titan's atmospheric aerosols is illustrated in their supplementary Fig. 6: a remarkable, detailed structure, consisting of an aromatic and a cyclohexane ring, connected and substituted by several linear and branched aliphatic chains, bearing one amino, two imino and two nitrile groups. Even though this is designated as a "probable" structure, it is not justified on the basis of the authors' dubious evidence for the presence of NH₃ and HCN.

Compounds of this type would pyrolyse to small unsaturated aliphatic and aromatic molecules. Benzene and its homologues are easy to detect by mass spectrometry at very low levels owing to their aromaticity. They give rise to abundant molecular ions at *m/z* 78 and 78 + 14*n*, respectively, which is well within the mass range (*m/z* 2–142) of the mass spectrometer aboard the Huygens probe³. To explain their absence, Israël *et al.* suggest that the spectrometer has low sensitivity above *m/z* 50.

To justify the extrapolation from NH₃ and HCN to complex organic matter of the type shown in their supplementary Fig. 6, Israël *et al.*² rely on laboratory analogues (tholins) of Titan's aerosols. These have been produced

in various laboratories ever since data from Voyager's 1980 fly-by of Titan showed that its atmosphere consists of nitrogen and a few per cent of methane. Irradiation of such mixtures with various energy sources invariably produced some amorphous materials^{1,4–6}, but these were only crudely characterized and not even partially separated for structural analysis. In support of their assertion that such a tholin on pyrolysis produces only NH₃ and HCN (but nothing else), the authors show a gas chromatogram in their supplementary Fig. 1. However, that figure shows only the region from about 2–10 min of the 60-min chromatogram, cutting off just after the emergence of HCN at 9.35 min, beyond which many other pyrolysis products are likely to emerge^{1,5}. The claim to have detected complex organic matter in Titan's atmospheric haze is therefore further undermined.

Other data obtained from the successful Cassini–Huygens mission contradict or correct some previously held notions about Titan, its origin and its environment^{3,7,8}. The idea that the observed haze must be due to aerosol particles consisting of large organic molecules of complex structure needs to be re-evaluated. The chemical nature of the haze on Titan still remains shrouded in mystery.

K. Biemann

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
e-mail: kbiemann@mit.edu

1. Israël, G. *et al.* *Space Sci. Rev.* **104**, 435–466 (2002).
2. Israël, G. *et al.* *Nature* **438**, 796–799 (2005).
3. Niemann, H. *et al.* *Nature* **438**, 779–784 (2005).
4. Khare, B. N. *et al.* *Adv. Space Res.* **4**, 59–68 (1984).
5. Coll, P., Coscia, D., Gazeau, M.-C., Guez, L. & Raulin, F. *Orig. Life Evol. Biosph.* **28**, 195–213 (1997).
6. Somogyi, A., Oh, C.-H., Smith, M. A. & Lunine, J. I. *J. Am. Soc. Mass Spectrom.* **16**, 850–859 (2005).
7. Porco, C. C. *et al.* *Nature* **434**, 159–168 (2005).
8. Zarnecki, J. C. *et al.* *Nature* **438**, 792–795 (2005).

doi:10.1038/nature05417

ASTROCHEMISTRY

Israël *et al.* reply

Replying to: K. Biemann *Nature* **444**, doi: 10.1038/nature05417 (2006)

Biemann¹ calls into question our preliminary interpretations of our experimental results². A comparison of laboratory and flight measurements should settle the uncertainties he raises. In addition to evaluating instrumental

characteristics such as the 'piston effect', a technique we used for injecting oven-gas content for gas chromatography–mass spectrometry (GC–MS), such studies should enable a wide range of possible compositions for Titan's

aerosols to be investigated. Our aerosol collection and pyrolysis (ACP) measurements in Titan's atmosphere can then be revisited.

Biemann doubts that we were detecting NH₃ and HCN after the pyrolysis steps at 600 °C, as well as questioning our deductions. If the detection is valid, our inference is clear. The data obtained from temperature sensors in flight undoubtedly show that the oven heaters did work nominally, so any volatile material collected by ACP along with the haze particles must have been vaporized during the 250 °C

heating step, and — as indicated by the reading of the oven pressure sensors — evacuated from the ACP–GC–MS. Consequently, only the refractory part of the aerosols could have remained in the oven afterwards. The detection of NH₃ and HCN in the products formed during the pyrolysis of this unknown, but refractory, material indicates that it must comprise a mixture of carbon-, hydrogen- and nitrogen-containing species.

Biemann suggests that the laboratory analogues of Titan's aerosols were only crudely characterized and not even partially separated for structural analysis; however, this is not correct. The laboratory tholins were recovered and studied using different techniques in order to determine their chemical composition and their cracking patterns, which are the parameters required for interpreting observational data^{3–7}. He also misinterprets our comments on the specific structure of the aerosols. Our supplementary Fig. 6 showing the probable structure of Titan's aerosols (ref. 2) is intended to illustrate a representative example of what could be part of the structure of the complex organic material, and not what it actually is. This structure was constructed from the nature of the identified pyrolysates: nitrile groups to produce HCN, and amino or imino groups to produce NH₃. This

type of illustration has previously been used⁸, based on experimental work⁹.

An important puzzle that will be addressed by laboratory simulations is why there are no fragments of high relative molecular mass in the spectra of the pyrolysate formed at 600 °C, which may be explained by their destruction during transport to the detector. We must, of course, accept the possibility that we could have been misled in our preliminary interpretations by a remarkable set of coincidences, which is why we are now completing a rigorous set of laboratory calibrations before reaching a final judgement on the results of the experiment.

G. Israël¹, C. Szopa¹, F. Raulin², M. Cabane¹, H. B. Niemann³, S. K. Atreya⁴, S. J. Bauer⁵, J.-F. Brun¹, E. Chassefière¹, P. Coll², E. Condé⁶, D. Coscia², A. Hauchecorne¹, P. Millian⁷, M. J. Nguyen², T. Owen⁸, W. Riedler⁹, R. E. Samuelson¹⁰, J.-M. Siguier⁷, M. Steller¹¹, R. Sternberg², C. Vidal-Madjar¹²

¹Service d'Aéronomie UMR 7620 CNRS, ISPL, Universities Paris 6 and Versailles-Saint Quentin, Verrières-le-Buisson 91371, France

²Laboratoire Interuniversitaire des Systèmes Atmosphériques, UMR 7583 CNRS, Universities Paris 12 and Paris 7, Créteil 94010, France

³NASA Goddard Space Flight Centre, Greenbelt, Maryland 20771, USA

⁴Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

⁵Institute for Meteorology and Geophysics, University of Graz, Graz 8010, Austria

⁶Centre National d'Etudes Spatiales (CNES), Toulouse 31401, France

⁷Office National d'Etudes et de Recherches Aérospatiales, Toulouse 31055, France

⁸Institute for Astronomy, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

⁹Space Research Institute, Graz 8010, Austria

¹⁰Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

¹¹Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz 8042, Austria

¹²Laboratoire de Recherche sur les Polymères, UMR 7581 CNRS, Thiais 94320, France

1. Biemann, K. *Nature* **444**, doi:10.1038/nature05417 (2006).
2. Israël, G. et al. *Nature* **438**, 796–799 (2005).
3. Khare, B. N. et al. *Icarus* **60**, 127–137 (1984).
4. McDonald, G. D. et al. *Icarus* **108**, 137–145 (1994).
5. Coll, P. et al. *Planet. Space Sci.* **47**, 1331–1340 (1999).
6. Ramirez, S. I. et al. *Icarus* **156**, 515–529 (2002).
7. Bernard, J. M. et al. *Icarus* **185**, 301–307 (2006).
8. Thompson, W. R., Henry, T. J., Schwartz, J. M., Khare, B. N. & Sagan, C. *Origins of Life* **19**, 475–476 (1989).
9. Lichtin, D. A. & Lin, M. C. *Chem. Phys.* **104**, 325–330 (1986).

doi:10.1038/nature05418

FRACTAL ANALYSIS

Revisiting Pollock's drip paintings

Arising from: R. P. Taylor, A. P. Micolich and D. Jonas *Nature* 399, 422 (1999)

We investigate the contentions that Jackson Pollock's drip paintings are fractals produced by the artist's Lévy distributed motion and that fractal analysis may be used to authenticate works of uncertain provenance^{1–5}. We find that the paintings exhibit fractal characteristics over too small a range to be usefully considered as fractal; their limited fractal characteristics are easily generated without Lévy motion, both by freehand drawing and gaussian random motion. Several problems must therefore be addressed before fractal analysis can be used to authenticate paintings¹.

An image is considered fractal if, when covered with a grid of square boxes of size L (refs 1, 6), the number of filled boxes $N \propto L^D$, where D , the fractal dimension, is a non-integer. It is generally accepted^{7,8} that, to establish power-law behaviour, the range of box sizes must span more than one or two orders of magnitude. For a Pollock drip painting, the range is typically limited to three orders of magnitude by the ratio of the canvas size to the size of the smallest feature. For all Pollock paintings examined so far, the box-counting curve, a plot of N versus L , is found to be a broken power law⁵ ($N \propto L^{D_L}$ for $L > l_T$, and $N \propto L^{D_D}$ for $L < l_T$, where l_T is a characteristic length). Thus, there are two power laws and less than two orders of magnitude to establish each. Some pitfalls of inferring fractal behaviour from such a limited range are demonstrated below.

For multicoloured, multilayered drip paintings (a large part of Pollock's oeuvre), it has been claimed⁴ that the visible part of each separate layer and the composite are all fractals. This inference is probably an artefact of the limited range of data used, because it is not a mathematical possibility if a model in which the individual layers are ideal fractals is considered⁶ — as, for example, in Cantor dusts or Sierpinski carpets. Calculation shows that the box-counting curves for the visible parts of each layer and the composite painting are not power laws, even though the individual layers are perfect fractals (Fig. 1).

Taylor *et al.* claim that the defining visual character of Pollock's drip paintings is their fractal nature⁴ and that fractals arise from the specific pouring technique developed by Pollock⁹. To test this, we drew a number of freehand sketches: Fig. 2 shows one we dub *Untitled 5*, and its box-counting curve. By the criteria espoused by Taylor and colleagues^{2–5}, it is a high-quality fractal; however, Fig. 2 undermines their claim^{2–5,9} that Pollock's works derive their artistic merit from their (limited) fractal content.

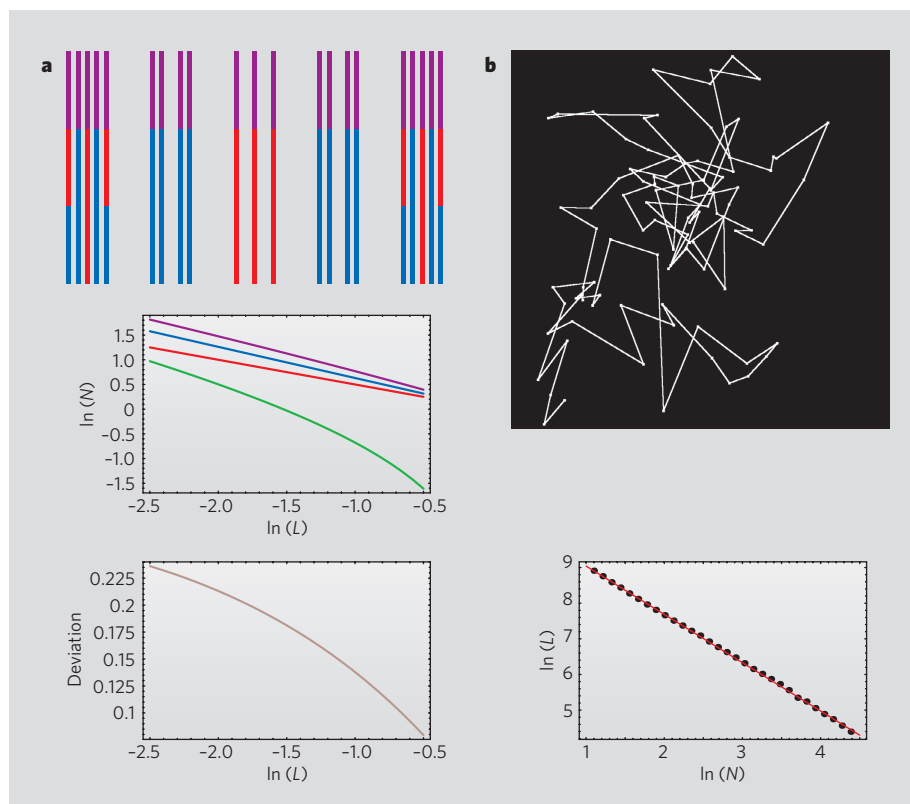


Figure 1 | Fractal barcode and gaussian walk. **a**, Top, a middle-third Cantor dust anchor layer (blue), overlaid with a second Cantor dust (red). Half-blue/half-red bars correspond to the intersection of the dusts; purple corresponds to their union. Centre, box-counting curves: blue dust (shown in blue), red dust (red), the uncovered part of the blue dust (green) and the composite (purple). The curvature of the traces indicates that the uncovered portion of the blue layer and the composite are not true fractals. To highlight the curvature of the composite, the lower graph shows the difference of the (rigorously linear) blue and purple curves. **b**, A linear fit to the box-counting curve of a 100-step gaussian walk has a slope of 1.35, with standard deviation $\chi = 0.025$.

Methods. For **a**, the blue dust is obtained by repeatedly dividing segments into three parts and retaining only the first and third; the red dust is obtained by dividing into nine parts and retaining the first, fifth and ninth parts. The 'fractal barcode' shows the appearance of the dusts after four iterations. For **b**, step size is calculated as $0.09 \times$ frame width. Smallest box size, 3 pixels. Sizes range over 1.4 orders of magnitude, with magnification $C = 1.12$ (see Fig. 2 for definition).

There are insufficient data^{10,11} to determine directly whether Pollock's motion while painting constituted a Lévy flight. Lévy motion has been inferred^{2–5} from the known result that Lévy flights leave a fractal trail⁶. However, the box-counting data of a simulated gaussian random walk of 100 steps, examined over only one-and-a-half orders of magnitude, are well described as a fractal ($D = 1.35$ in our simulation; Fig. 1) — although sufficiently long gaussian random walks are rigorously known to be non-fractal, with dimension $D = 2$. Thus, the limited fractal content of Pollock's work does not require Lévy flights for its explanation. Furthermore, Lévy motion has no natural length scales, whereas Pollock's paintings and

motion seem to have many⁴.

Box-counting authentication techniques assume that the parameters, l_T , D_D and D_L and χ_2 (standard deviation; Fig. 2, legend) show characteristic trends that distinguish Pollock's paintings from those of his imitators^{5,9}. For multifractals¹² and non-fractals, χ_2 is not an intrinsic characteristic of the image but depends on the magnification, C (Fig. 2, legend). If Pollock's paintings are multifractal¹³, it would be imprudent to use χ_2 as a characteristic parameter.

Also, we question the presumption that parameters such as l_T for Pollock's works are essentially random variables, circumscribed by a range determined⁹ by examination of just 17

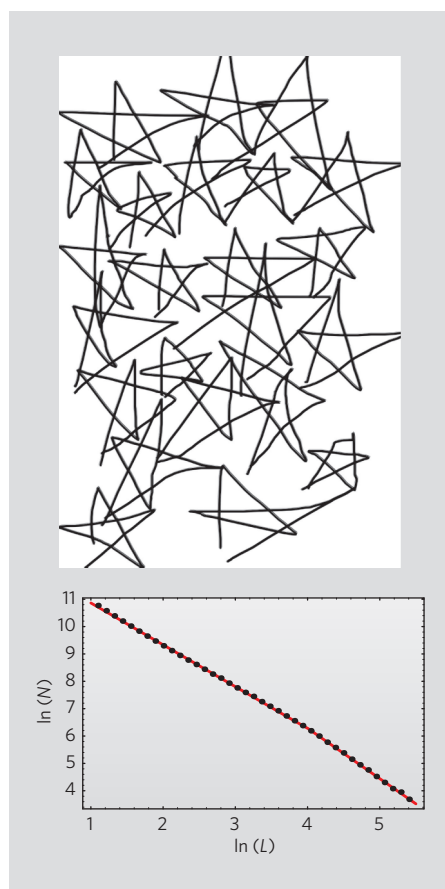


Figure 2 | *Untitled 5* and its box-counting curve. The best broken power-law fit to these data corresponds to slopes of $D_D = 1.53$ and $D_L = 1.84$. The break occurs at $\ln(L) \approx 4$; the standard deviation of the data from the fit, χ , is 0.022. *Untitled 5* (top) fulfils all the criteria used in box-counting authentication that have been made public: it shows a broken power-law behaviour with $D_D < D_L$ and, for a magnification factor C similar to that used by Taylor *et al.*, has a χ value in the 'permissible' range of $0.009 < \chi < 0.025$.

Methods. All our sketches, including *Untitled 5*, are freehand drawings made in Adobe Photoshop using a 14-point Adobe Photoshop 'paintbrush'. The paintbrush leaves a mark when dragged continuously across the 'canvas' by a computer mouse. Although not drip paintings, these patterns are human- and not computer-generated, in the sense that these terms are commonly understood and as used by Taylor *et al.*⁴. The box sizes were chosen to be hC^n , where $n = 1, 2, 3, \dots, M$. Here $h = 3$ pixels is the smallest box size, the magnification factor $C = 1.12$, the cutoff M determines the biggest box size, and the box sizes range over about 1.9 orders of magnitude.

drip paintings (out of about 180). Further study of systematic effects is needed. As the values measured by Taylor *et al.*⁹ are not available, a consensus on the limiting range, if there is one, can emerge only after other groups have replicated and extended the box-counting data set. Finally, we note that *Untitled 5* fulfils all the criteria used in box-counting authentication that have so far been made public.

Katherine Jones-Smith, Harsh Mathur

Department of Physics, Case Western Reserve

University, Cleveland, Ohio 44106, USA

e-mail: hxm7@case.edu

1. Abbott, A. *Nature* **439**, 648–650 (2006).
2. Taylor, R. P., Micolich, A. P. & Jonas, D. *Nature* **399**, 422 (1999).
3. Taylor, R. P., Micolich, A. P. & Jonas, D. *Phys. World* **12**, 25–28 (1999).
4. Taylor, R. P., Micolich, A. P. & Jonas, D. *Leonardo* **35**, 203–207 (2002).
5. Taylor, R. P. in *Art and Complexity* (eds Casti, J. & Karlqvist, A.) 117–144 (Elsevier, Amsterdam, 2003).
6. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1977).

7. Avnir, D., Biham, O., Lidar, D. & Malcai, O. *Science* **279**, 39–40 (1998).
8. Mandelbrot, B. B. *Science* **279**, 783 (1998).
9. Taylor, R. P. *et al. Patt. Recog. Lett.* (in the press) (materials.science.uoregon.edu/taylor/art/TaylorPRL.pdf).
10. Namuth, H. & Falkenberg, P. *Jackson Pollock* (Collectors Series film, The Museum of Modern Art, New York, 1950).
11. Landau, E. G. *Jackson Pollock* (Thames and Hudson, London, 1989).
12. Ott, E. *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 2002).
13. Mureika, J. R., Cupchik, G. C. & Dyer, C. C. *Leonardo* **37**, 53–56 (2004).

doi:10.1038/nature05398

FRactal Analysis

Taylor et al. reply

Replying to: K. Jones-Smith & H. Mathur *Nature* 444, doi:10.1038/nature05398

Our use¹ of the term 'fractal'² is consistent with that by the research community. In dismissing Pollock's fractals^{1,3} because of their limited magnification range, Jones-Smith and Mathur⁴ would also dismiss half the published investigations of physical fractals⁵. On the basis of previous debates on limited-range fractals^{5,6}, a fractal description is particularly appropriate for Pollock's patterns because it is physically reasonable and because it is useful for condensing the description of a complex geometry, as we now describe.

Fractal description is physically reasonable because Pollock's technique involved a motion-dominated process at large length scales and a paint-dominated process at small scales, with a transition between the two expected at several centimetres. The fractal behaviours observed in the scaling plots above and below this transition are physically reasonable — many physiological processes⁷, including human motion^{8,9}, are fractal, and falling liquid can itself be fractal¹⁰.

However, for *Untitled 5* there is no physical reason to expect a transition between two processes at $\ln(L) \approx 4$. In the absence of this artificial transition, the scaling plot does not fit fractal behaviour. Furthermore, when we generate 'drawings' similar to *Untitled 5*, some plots curve down and some curve up (Fig. 1a), in contrast to the consistent behaviour observed in Pollock's poured paintings. Combined, these observations demonstrate that the "freehand sketches" of Jones-Smith and Mathur⁴ are not fractal. Hence, *Untitled 5* does not undermine any claim that we have made about Pollock's work.

As to the usefulness of a fractal description, Pollock's fractals satisfy all the criteria for usefulness proposed for limited-range fractals⁵. Furthermore, our fractal description of Pollock's patterns is useful across disciplines, for example in psychology experiments (showing that Pollock's fractals induce the same perceptual responses as mathematical fractals¹¹), and

in applications of fractal geometry to distinguish paintings by different artists^{3,12}.

We disagree with the claim by Jones-Smith and Mathur⁴ that Pollock's fractal characteristics are easily generated by gaussian random motion. Our simulations generate scaling plots with variable curvature (Fig. 1b), in contrast to the consistent behaviour shown in Pollock's paintings¹².

It is both mathematically and physically possible for two exposed patterns and their composite all to be fractal. This depends on the relative densities of the exposed patterns and the scaling behaviour of boxes containing both patterns. Jones-Smith and Mathur's Cantor dust does not apply to Pollock paintings, where the overlap of layers is considerably more complicated. Furthermore, their Fig. 1a features box counts of less than one (in reality, a box is either filled or not). The gradient of their green trace becomes greater than one (impossible for an object with $0 < D < 1$). Their Fig. 1a also misleads because it concentrates on large scales — the deviation of the blue and composite box counts becomes insignificant at finer scales (Fig. 1c).

We performed an actual box count on the dust analysed by Jones-Smith and Mathur (Fig. 1d) in order to show that their analytical simulations are flawed. Their expressions are evaluated at each iteration, instead of correctly box counting after fully generating the pattern. Their green curve is generated by using an expression that holds only for specific cases in which a 'box' exactly matches the width of a Cantor dust segment: it is therefore invalid for most of the continuous range the authors display.

Finally, we keep the value of the magnification factor C constant in recent authenticity procedures because χ is dependent on C (ref. 12, which also gives our authenticity criteria, which are not confidential). We encourage further research.

R. P. Taylor^{*}, A. P. Micolich[†], D. Jonas[‡]

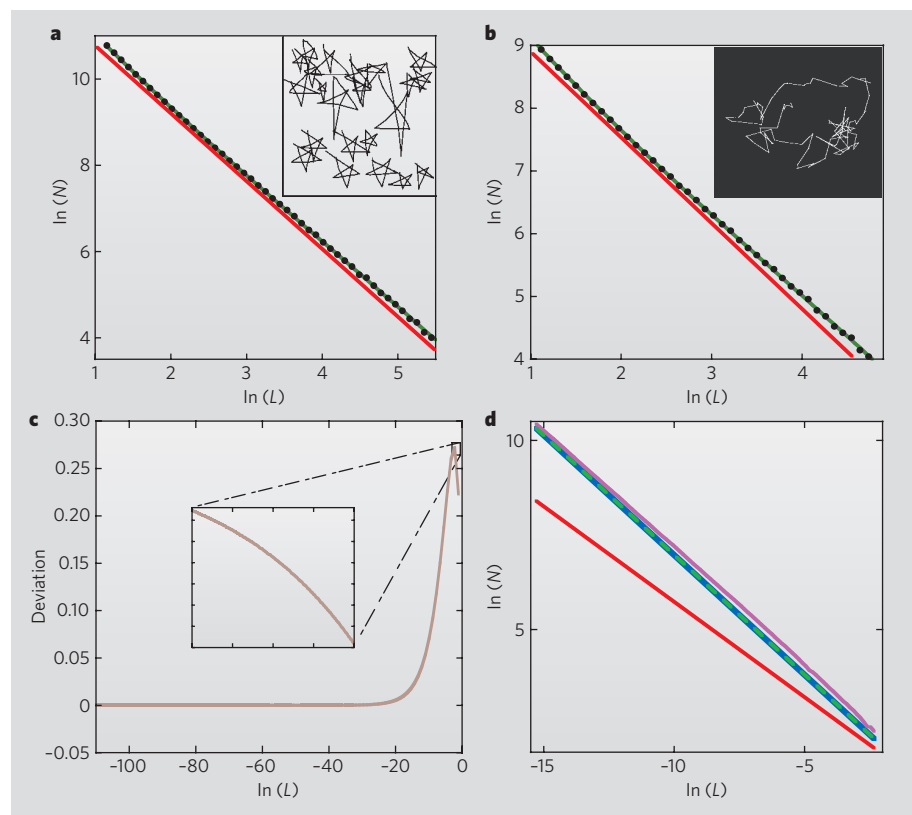


Figure 1 | Analysis of freehand, gaussian and Cantor dust patterns. **a, b,** Example images (insets) and scaling plots of **a**, 'freehand' and **b**, gaussian patterns. Both plots deviate from the scaling behaviour found in Pollock's paintings. **c**, Brown trace from Fig. 1a of Jones-Smith and Mathur⁴, plotted over a large scaling range and showing that the deviation is insignificant over most of the graph. Inset, limited region selected by Jones-Smith and Mathur⁴. **d**, A standard box count performed on the Cantor dust of ref. 4 (see their Fig. 1a for colour coding). In contrast to their analytical simulations, our green trace is rigorously linear over six orders of magnitude, deviating from the blue trace only when the box size matches the dust's finest features. T. P. Martin, B. C. Scannell and M. S. Fairbanks contributed to this analysis.

Methods. For **a, b**, the 'freehand sketch' and gaussian random walk were generated as in Jones-Smith and Mathur⁴. Green lines are third-order polynomial fits to the data; red lines act as a guide to the eye, illustrating the data's upward curvature. For **c**, we iterate the analytical expression of Jones-Smith and Mathur⁴ to produce our brown curve. For **d**, we generate their Cantor dust to seven iterations to ensure appropriate statistics and then perform a standard box count. To ensure sufficient counting statistics, the largest box size plotted corresponds to one-tenth of the image.

*Physics Department, University of Oregon,
Eugene, Oregon 97403, USA

e-mail: rpt@uoregon.edu

†School of Physics, University of New South
Wales, Sydney, New South Wales 2052, Australia

1. Taylor, R. P., Micolich, A. P. & Jonas, D. *Nature* **399**, 422 (1999).
2. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1977).
3. Mureika, J. R., Dyer, C. C. & Cupchik, G. C. *Phys. Rev. E* **72**, 046101-1-15 (2005).
4. Jones-Smith, K. & Mathur, H. *Nature*

doi:10.1038/nature05398 (2006).

5. Avnir, D., Biham, O., Lidar, D. & Malcai, O. *Science* **279**, 39-40 (1998).
6. Mandelbrot, B. B. *Science* **279**, 783 (1998).
7. Bassingthwaite, J. B., Liebovitch, L. S. & West, B. J. *Fractal Physiology* (Oxford Univ. Press, 1994).
8. Hausdorff, J. M. et al. *J. Appl. Physiol.* **80**, 1448-1457 (1996).
9. Doyle, T. et al. *Int. J. Med. Sci.* **1**, 11-20 (2004).
10. Shi, X. D. et al. *Science* **265**, 219-222 (1994).
11. Aks, D. & Sprott, J. C. *Empir. Stud. Arts* **14**, 1-16 (1996).
12. Taylor, R. P. et al. *Patt. Recog. Lett.* (in the press).

doi:10.1038/nature05399

Protein delivery into eukaryotic cells by type III secretion machines

Jorge E. Galán¹ & Hans Wolf-Watz²

Bacteria that have sustained long-standing close associations with eukaryotic hosts have evolved specific adaptations to survive and replicate in this environment. Perhaps one of the most remarkable of those adaptations is the type III secretion system (T3SS)—a bacterial organelle that has specifically evolved to deliver bacterial proteins into eukaryotic cells. Although originally identified in a handful of pathogenic bacteria, T3SSs are encoded by a large number of bacterial species that are symbiotic or pathogenic for humans, other animals including insects or nematodes, and plants. The study of these systems is leading to unique insights into not only organelle assembly and protein secretion but also mechanisms of symbiosis and pathogenesis.

Made up of more than 20 proteins, T3SSs are among the most complex protein secretion systems known in bacteria. Although the components of type III secretion machines are generally conserved, comparison of their amino acid sequence reveals the existence of different ‘clades’ of related T3SSs¹. Despite this conservation, the actual arsenal of bacterial proteins that they deliver (collectively known as ‘effectors’) is unique to each system. T3SSs are thought to be evolutionarily related to the bacterial flagellum, although this relationship has been the subject of some debate². In this article, however, we will discuss the most salient features of T3SSs, but we will not discuss the related flagellar system³. We have focused on what we believe are the general principles that govern the function of these biological machines. Readers can consult other reviews on this topic for more details on specific systems^{4,5}.

The assembly of a protein-delivery machine

A central component of T3SSs is a supramolecular structure known as the needle complex, which mediates the passage of the secreted proteins through the multi-membrane bacterial envelope. Although the needle complex was originally identified in *Salmonella typhimurium*⁶, it has been subsequently detected in several other bacteria and it is therefore believed to be a core component of all T3SSs^{7–9}. It is composed of a multi-ring base, which anchors the structure to the bacterial envelope, and a needle-like projection that protrudes several nanometres from the bacterial surface (Fig. 1). The base is traversed by the inner rod, which is a cylindrical substructure that connects the needle to the basal side of the base substructure. The entire needle complex is traversed by a narrow channel (~28 Å in diameter), which most probably serves as the conduit for proteins travelling through this secretion pathway. *Salmonella typhimurium* assembles heterogeneous needle complexes of varying symmetries, ranging from 19- to 22-fold, on the vertical axis¹⁰. However, this heterogeneity may not be functionally significant, but may simply reflect the intrinsic difficulty in maintaining ‘quality control’ during the assembly of such a many-fold symmetry structure.

Assembly of the needle complex proceeds in a stepwise manner, in which the assembly of the base substructure precedes the assembly of the inner rod and needle (Fig. 2; refs 11–13). In *S. typhimurium* the

base is composed of equimolar amounts of three proteins: InvG, a member of the secretin family of proteins that makes up the outer rings; and PrgH and PrgK, which are thought to form the rest of the structure¹⁰. Homologues of these proteins can be detected in other T3SSs, suggesting that the composition and assembly pathway are likely to be conserved. Proteins destined to make up the base are delivered to the bacterial envelope by the universal Sec protein secretion machinery, which is consistent with the observation that these proteins possess signal sequences for this common secretion pathway. In some T3SSs, assembly of the outer rings requires an accessory outer membrane protein that serves as a chaperone to presumably

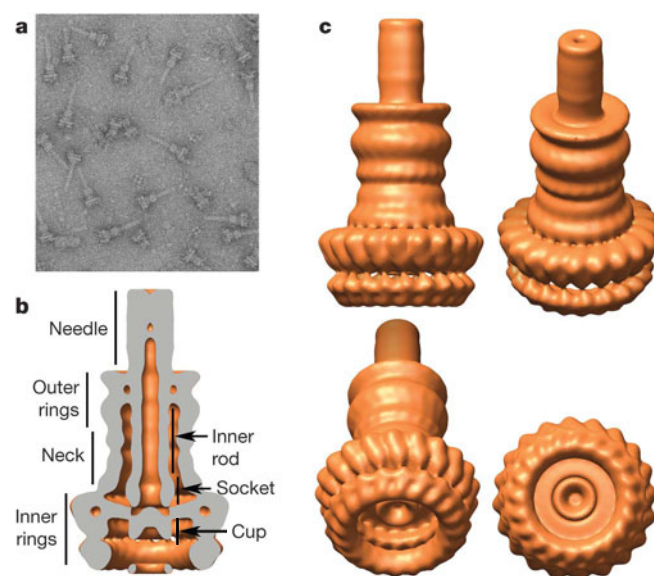


Figure 1 | Needle complex of *Salmonella typhimurium*. **a**, Electron micrographs of negatively stained isolated needle complexes. **b**, Cross-section of the structure of the needle complex indicating the location of its different substructures. **c**, Surface rendering of the structure of the needle complex. Shown here are different views of the structure of a 20-fold complex with 20-fold symmetry imposed.

¹Section of Microbial Pathogenesis, Yale University School of Medicine, Boyer Center for Molecular Medicine, New Haven, Connecticut 0636, USA. ²Department of Molecular Biology, Umeå University, Umeå SE-901 87, Sweden.

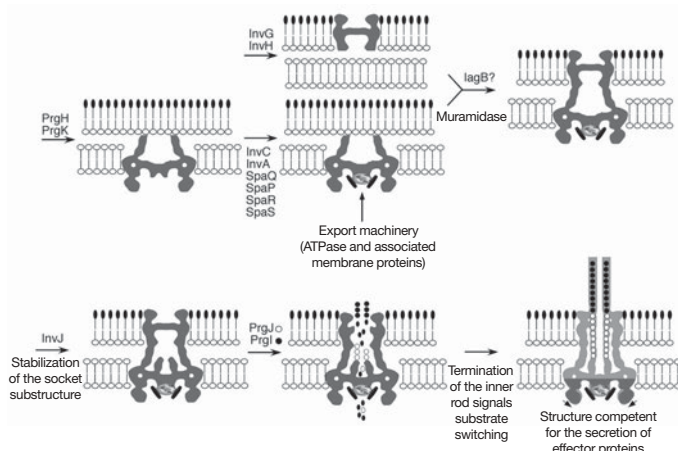


Figure 2 | Model of the assembly pathway of the needle complex.

Assembly of the base substructure occurs in discrete steps, and its completion is followed by the recruitment of accessory proteins including an ATPase (InvC) that is involved in the recognition and unfolding of substrates of the type III secretion machine. A muramidase (IagB) may help passage of the structure through the peptidoglycan layer. Completion of the base initiates secretion of needle (PrgI), inner rod (PrgJ) and regulatory (InvJ) proteins. InvJ stabilizes the socket substructure of the base to allow assembly of the inner rod. Termination of the inner rod results in conformational changes on the cytoplasmic side of the base, which leads to reprogramming of the secretion machine to begin the secretion of effector proteins. InvH chaperones the assembly of InvG to form the outer ring but does not form part of the final structure. InvA, SpaP, SpaQ, SpaR and SpaS are polytopic inner membrane proteins that form part of the 'export apparatus'. The nomenclature of relevant proteins is from *S. typhimurium*. See text for details.

facilitate the proper folding and multimeric arrangement of the outer ring subunit^{14,15}. Determination of the atomic structure of EscJ (the orthologue of *S. typhimurium* PrgK in enteropathogenic *Escherichia coli*) and the use of molecular modelling, indicate that this protein could form a large multi-subunit 'ring' that may serve as an initiator of needle complex assembly^{16,17}. However, more experiments will be required to validate this attractive hypothesis.

Assembly of the base substructure presumably leads to the recruitment of a subset of highly conserved, accessory, inner membrane proteins, which are thought to facilitate the passage of secreted proteins through the inner membrane. By analogy to the flagellar apparatus, it has been proposed that another compartment, the so-called 'C ring', may be assembled in the bacterial cytoplasm immediately beneath the needle complex¹⁸. In *Shigella* spp., this compartment may be formed at least in part by the T3SS protein Spa33, which was shown to be associated with the needle complex base¹⁹. However, homologues of Spa33 in other T3SSs are secreted into the culture supernatant²⁰, which would not be consistent with this proposed function. More studies will be required to ascertain the role of this protein in type III secretion and better define the putative substructure that it may form.

Once fully assembled, the base substructure begins to function as a 'type III protein secretion machine'. However, this submachine is exclusively devoted to the secretion of proteins that are components of, or necessary for the assembly of the inner rod and needle substructures. In *S. typhimurium* these proteins are PrgJ (the putative inner rod protein), PrgI (the needle protein) and InvJ (a regulatory protein that is required for appropriate and efficient assembly of the needle complex, but that does not form part of the final structure; see below)^{10,12}. On completion of the needle complex assembly, the secretion machine changes specificity and becomes competent for secreting effector proteins^{12,21}. Substrate switching and reprogramming of the export machinery is also observed during flagellar assem-

bly³. In this case, once the hook substructure reaches a certain length, the export machinery switches substrate specificity to export proteins necessary for the assembly of the flagellar filament. Substrate switching in the T3SS and flagellar systems is largely dependent on the function of a family of somewhat related accessory proteins (for example, InvJ (ref. 21) and YscP (ref. 22) in the case of the *Salmonella* spp. and *Yersinia* spp. T3SSs, respectively, and FliK (ref. 23) in the case of the flagellar hook). Mutant strains lacking these proteins assemble needle complexes with much longer needles or flagella with much longer hooks. The mechanisms by which these secretion machines change substrate specificity are incompletely understood. Comparison of the 17 Å resolution structure of the *S. typhimurium* fully assembled needle complex, which is competent for the secretion of effector proteins, with that of the base alone, which is competent for the secretion of the needle and the inner rod proteins, revealed intriguing differences that may provide the structural basis for this secretion reprogramming¹⁰. Very significant changes were observed on the cytoplasmic side of the base, which is the surface of the needle complex that would be available for interaction with the secreted proteins themselves or accessory proteins involved in substrate recognition (see below). In particular, a cup-like structure, situated at the centre of the cytoplasmic face of the base (see Fig. 1), undergoes a marked 'downward' movement on completion of the assembly of the needle substructure. Furthermore, the more distal inner ring, which is presumably located within the bacterial cytoplasm, undergoes a 'clumping' and 'downward' movement, further redefining the shape of the cavity that surrounds the area of the needle complex that is likely to be involved in substrate recognition. How these conformational changes affect substrate recognition, however, is not known. Furthermore, the mechanisms by which the InvJ, YscP and FliK proteins mediate substrate switching are very poorly understood.

At least three seemingly incompatible models to explain the mechanisms of substrate switching during the assembly of these organelles have been proposed. One model put forward for the flagellar system proposes that a flagellar substructure, the cytoplasmic C ring, acts as a 'measuring cup' that, once filled with hook proteins, permits their secretion and the assembly of a hook of defined length²⁴. When the C-ring is emptied, the FliK protein (perhaps in conjunction with other regulatory proteins, for example, FlhB) interacts with the secretion apparatus and triggers substrate switching. Although attractive, there is little mechanistic and experimental evidence in support of this hypothesis.

The second model, formed to account for the mechanisms of needle-length control in the *Yersinia* spp. T3SS needle complex, proposes that the YscP protein works as a 'molecular ruler'²². In this model, YscP is proposed to exert its function as an extended polypeptide that is anchored to both the tip of the growing needle and the base of the needle complex. The full extension of the YscP 'ruler' would signal that the needle has reached the appropriate length and would somehow convey the information to the secretion apparatus to switch substrates. In support of this model, lengthening or shortening YscP by inserting or removing sequences at a discrete domain resulted in longer or shorter needles. In all likelihood, the needle substructure grows by addition of subunits that travel through the central channel and are added at the distal growing tip²⁵. Therefore, in the context of this model it is not clear how these subunits could travel through the narrow central channel of the needle complex when it is occupied by the 'measuring ruler'.

A third model was proposed based on the observation that needle complexes isolated from the *S. typhimurium* $\Delta invJ$ mutant lack the inner rod substructure²⁶. This observation indicates that assembly of the needle and inner rod substructures can occur independently, although firm anchoring of the needle does require the presence of the inner rod because needles isolated from the $\Delta invJ$ mutant easily detach from the base. The high-resolution structure of the needle complex showed that in the wild type, the inner rod is anchored to

a 'socket-like' structure located on the basal side of the base¹⁰. This 'socket' is thought to serve as a symmetry adaptor between the helical inner rod and the *N*-fold symmetric base. The $\Delta invJ$ mutant lacks the socket²⁶, indicating that this structure may be required to nucleate the assembly of the inner rod, which is missing in this mutant. Furthermore, the cytoplasmic side of the base isolated from the $\Delta invJ$ mutant is very similar to that of the wild-type base before needle assembly, which is consistent with the fact that both structures have essentially identical substrate specificity. These observations led to the proposal of a model for the mechanism of substrate-specificity switching in which the completion of the inner rod would determine the timing of substrate switching (and hence needle length)²⁶. In this model, the termination of the inner rod and the firm anchoring of the needle would lead to conformational changes on the cytoplasmic side of the base, resulting in the reprogramming of the secretion apparatus and rendering it competent for the secretion of effector proteins. Consistent with this model, changes in the relative concentrations of PrgJ and PrgI, which would presumably lead to relatively faster or slower assembly of the inner rod, led to needles of different lengths (that is, shorter in the case of overexpression of PrgJ and longer in the case of overexpression of PrgI). In the context of this model, the InvJ protein would control the assembly of the inner rod by stabilizing a conformation of the socket of the needle complex base that is permissive for the anchoring of the inner rod, in a manner similar to that proposed for the flagellar 'capping' proteins. These flagellar proteins, which like InvJ control flagellar assembly but are not part of the final structure and are discarded into the culture supernatant, are thought to work as scaffolds or 'kinetic traps' to facilitate the nucleation or anchoring of different flagellar substructures³. More studies will be required to reconcile these different models. In any case, it is clear that these models are incomplete because they do not account for the function of other accessory proteins (for example, the flagellar FlhB or its T3SS homologues) known to be important for substrate switching²⁷. It is also evident that it is important for bacteria to assemble needles of a minimum length so that other bacterial surface structures do not interfere with type III secretion function^{28,29}.

Multiple signals to ensure substrate specificity

Type III secretion machines must be able to select the rather small number of substrate proteins that are destined to travel this pathway. In addition, some bacteria encode more than one T3SS, which are

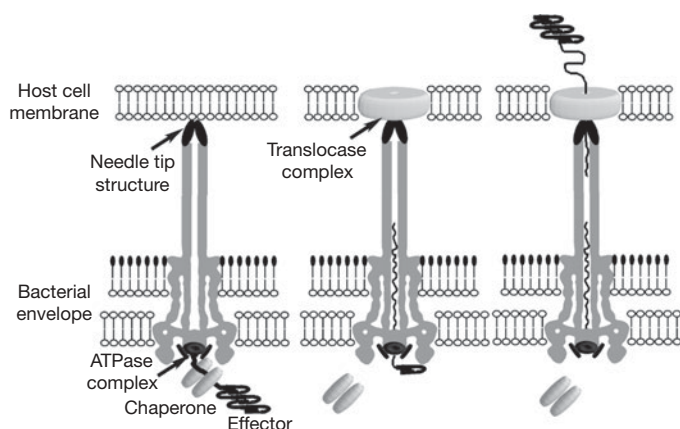


Figure 3 | Model for substrate recognition and delivery of proteins by type III secretion machines. The effector–chaperone complex is recognized by the secretion machinery, including a type-III-secretion-associated ATPase. The ATPase 'strips' the chaperone from the complex, which remains within the bacterial cell, and mediates the unfolding and 'threading' of the effector protein through the central channel of the needle complex. A 'translocator complex' made up of proteins also secreted by the T3SS is assembled on the host cell membrane and mediates the passage of the effector proteins through the target cell membrane. The translocated effectors re-fold within the host cell to carry out their function. See text for details.

often expressed simultaneously. Therefore, the mechanisms of substrate recognition must warrant a level of specificity capable of ensuring that the correct substrates are targeted to the appropriate machine. The narrow opening of the channel that traverses the needle complex and serves as the conduit for the type III secreted proteins also dictates that, once recognized by the secretion machine, substrates must be unfolded before their secretion. Furthermore, evidence is beginning to mount indicating that the secretion process follows a hierarchy with a predetermined order in which different proteins are engaged and secreted by these machines^{21,30,31}. Therefore, it is not surprising that the mechanisms of substrate recognition are complex, involving multiple signals and accessory proteins³² (Fig. 3).

Most, if not all, type III secreted proteins possess a secretion signal located within the first ~20–30 amino acids^{33,34}. Unlike Sec-dependent signal sequences, however, T3S signals are not cleaved on secretion. In general, these signals do not seem to have conserved features that are shared even among substrates of the same type III secretion machine. Many studies have even shown that introduction of frame-shift mutations within the secretion signals of at least some type III secreted proteins does not prevent their secretion. In fact, this observation led to the proposal that the 'secretion signal' may be located in the coding messenger RNA and not in the polypeptide³⁵. However, persuasive evidence has been presented that, at least in some type III secreted proteins, the secretion signal does indeed reside within the amino acid sequence and not the mRNA^{36–38}.

How can the demonstrated specificity of these secretion machines be reconciled with the obvious tolerance for change in the secretion signal? Although the answer to this question is still a matter of debate, it is conceivable that the following considerations may provide the basis for an explanation. First, it is possible that the presence of non-structured flexible segments at the amino terminus may be all that is required for a sequence to function as a type III secretion signal. However, it is unlikely that this feature alone could ensure specificity, as it is likely that many proteins that are not substrates of these machines exhibit this feature. Therefore other elements must ensure the specificity of substrate selection. Second, an additional layer of specificity may be conferred by accessory proteins such as a family of customized cytosolic chaperones that specifically bind at least some of the type III secreted proteins³⁹. These type-III-secretion-associated chaperones are small, acidic, dimeric proteins, which unlike other chaperones, lack ATP-binding or ATP-hydrolysing activities⁴⁰. Although in general type-III-secretion-associated chaperones do not

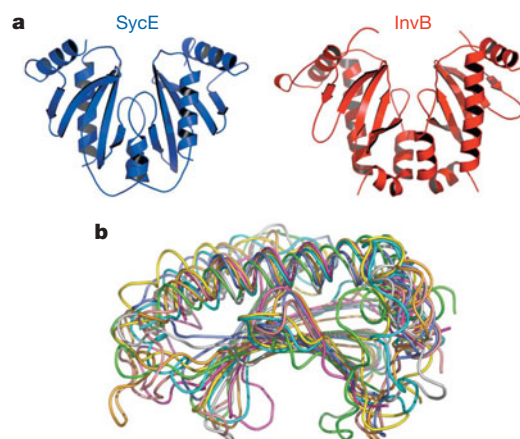


Figure 4 | Crystal structures of type III secretion chaperones. **a**, Ribbon representation of the crystal structures of type III secretion chaperones SycE from *Yersinia pseudotuberculosis* and InvB from *Salmonella typhimurium*, which bind to one of multiple effector proteins. **b**, Structural alignment of multiple type III secretion chaperones. The structures of AvrPphF (green), InvB (cyan), SigE (magenta), Spa15 (yellow), SicP (grey), SycH (purple) and YscB (orange) are shown.

share significant primary amino acid sequence similarities among themselves, the crystal structures of several of these chaperones have clearly shown that they are highly related molecules⁴¹ (Fig. 4). In general, these chaperones bind a ~50–100 amino acid domain of the secreted protein, located immediately downstream from the N-terminal secretion signal^{34,42}. Although most often a given chaperone is specific for a single secreted protein, there are chaperones that can bind multiple target proteins. The co-crystal structures of several of these chaperones bound to their cognate secreted protein have provided very valuable insight into their function^{43–45}. These structures showed that these chaperones maintain the chaperone-binding domain of their cognate secreted proteins in a non-globular conformation that nevertheless maintains secondary structure. This observation has led to the proposal that at least one of the functions of these chaperones must be to ‘prime’ the secreted proteins for rapid unfolding before secretion⁴¹. Another proposed function is to prevent the premature and undesirable interactions of the cognate secreted protein with other components of the type III secretion machinery within the bacterial cell⁴². In addition, it is clear that these chaperones play a key role in targeting the secreted protein to the cognate type III secretion apparatus. Consistent with this hypothesis, removal of the chaperone-binding domain of at least some type III secreted proteins prevents their secretion through their associated T3SS. Intriguingly, in some cases these deletion mutants are targeted to the flagellar export apparatus^{38,45}. These results indicate that the N-terminal signal sequence may be a ‘generic’ type III secretion signal, but other sequences and accessory proteins such as the type-III-secretion-associated chaperones may confer the specificity.

It has been suggested that the overall three-dimensional path followed by the chaperone-binding domain when associated to its cognate chaperone could serve as a conserved ‘universal’ signal for recognition by T3SSs⁴⁴. However, this is unlikely as comparison of the available co-crystal structures of several chaperone–effector complexes indicates that the polypeptide paths are too different to serve as universal signals⁴⁵. It is therefore more probable that other structural features common to all chaperones may serve as recognition signals. Given the high degree of three-dimensional structural relatedness among these chaperones, this possibility seems likely. In addition a common feature has been identified in the chaperone-binding domain of effector proteins⁴⁵. However, although this motif seems to be important for targeting the effector proteins to their cognate chaperones, there is no evidence to support a role for this domain in substrate recognition by the secretion apparatus.

In addition to recognizing secretion signals on the chaperone–effector complex, the secretion machine must ‘strip’ the chaperone from the effector protein because type-III-secretion-associated chaperones remain in the bacterial cytosol after delivery of the effector proteins to the secretion apparatus. The chaperone–secreted-protein interface is substantial (up to 6,200 Å² in some cases), therefore this presents a significant challenge for the secretion machine. Furthermore, the effector domains of the secreted protein, which are usually located carboxy-terminal to the chaperone-binding domain, are indeed folded when bound to the chaperone⁴⁶. The limitation in size of the secretion channel (estimated to be ~28 Å) dictates that this domain must be unfolded before secretion. Recently, it has been shown that a highly conserved ATPase associated with the type III secretion apparatus has a critical role in all these steps. *In vitro* studies showed that the ATPase binds both the chaperones as well as the chaperone–effector complexes^{47,48}. More importantly, addition of the ATPase to the chaperone–effector complex *in vitro* results in the dissociation of the complex and the unfolding of the effector domain of the effector protein⁴⁸. Therefore the T3SS-associated ATPases play a critical role in the recruitment and unfolding of the T3SS substrates, a function consistent with the structure and proposed localization of the ATPases in close proximity to the secretion machine⁴⁹. Furthermore, this unfolding activity may be critical for energizing the secretion process. These observations

suggest intriguing parallels between T3SS-associated ATPases and AAA⁺ ATPases. The members of the AAA⁺ ATPase protein family are involved in a large variety of biological processes and they exert their function by dissociating protein complexes or unfolding specific protein substrates⁵⁰. More remarkably, the similarity may extend to some essential features of the substrate recognition process. For example, AAA⁺ machines, such as the ClpA/ClpP protease, select substrates by recognizing poorly conserved unstructured short peptide signals often located at the N- or C-terminus of the target proteins. In addition, substrate specificity is often ensured through the activity of accessory proteins. Therefore type III secretion and the translocation of substrates into triple AAA⁺ machines may share some mechanistic features.

Protein delivery into target cells is not a needle stick

Addition of type-III-secreted effector proteins to the culture medium is not sufficient to promote their translocation into cultured host cells and the bacterial ‘delivery device’ is absolutely essential for protein translocation. However, the mechanism by which type III secretion machines carry out this function remains one of the least understood aspects of type III secretion. A model was proposed in which the needle itself, through the activity of the needle protein, would ‘punch’ a hole in the host cell membrane placing its ‘tip’ inside the cell and thereby delivering the proteins into the target cell cytosol⁵¹. Attractive as this model may seem, it is clearly incorrect. Indeed, the needle complex alone is not capable of mediating protein injection but needs the activity of a subset of somewhat conserved proteins that are themselves secreted by the T3SS^{52,53}. This group of proteins (known as ‘translocators’) are thought to insert in the target cell membrane forming a channel through which the effector proteins can pass on their way to the target cell cytosol^{53,54}. Consistent with this model, strains lacking the translocators are able to secrete the effector protein but are unable to deliver them into the target cells^{52,53}. These mutants can assemble the needle complex, indicating that the needle itself is insufficient to mediate translocation. A possible scenario is that the needle actually ‘docks’ onto the pore or channel made up by the translocators thereby allowing the direct delivery of effector proteins into the target cell (Fig. 3). Such docking may be facilitated by accessory structures located at the tip of the needle. One of these structures, recently identified in *Yersinia enterocolitica*, is formed by a single protein, LcrV (ref. 55). Another more complex structure has been visualized in the T3SS of enteropathogenic *E. coli* and some plant pathogenic bacteria^{56,57}. This structure, which is also formed by a single protein (for example, EspA in the case of the *E. coli* T3SS), takes the form of a long appendage that extends from the tip of the needle and presumably serves as a ‘bridge’ linking the needle with the bacterial translocators on the target cell membrane. It is not known how the insertion of the translocators in the target cell plasma membrane, the assembly of the accessory structures at the tip of the needle, the putative docking of the secretion machine to the translocator channel, and the actual delivery of the effector proteins are ultimately coordinated.

Sensing the target to deliver goods at the right time

T3SSs are highly regulated to ensure that they function at the appropriate time. In their simplest form, the regulatory mechanisms ensure that the secretion machine is deployed in the bacterial envelope when the appropriate cues are present. These regulatory mechanisms are largely transcriptional and rather specific for each T3SS (ref. 58). In addition, in a subset of T3SSs there are regulatory mechanisms to ensure that the secretion machine is activated only on bacterial contact with the target cell^{30,59}. This regulatory mechanism is presumably in place to prevent the premature, non-productive release of the effector proteins from the cell. Although this activation step can be triggered *in vitro* with some specific treatments (for example, lowering the calcium concentration or adding Congo red triggers the *Yersinia* spp. and *Shigella* spp. T3SSs, respectively), the physiological

signals must derive from target cell contact. The nature of the stimulating signal is unknown, but it is unlikely to involve a soluble factor because it is difficult to envisage how such a stimulant could ensure that the activation occurs only on bacterial cell contact. More likely, the bacteria or the secretion machine must be able to sense contact with the cell, leading to the activation of the secretion apparatus.

How could this mechanism work? An attractive hypothesis is that the needle, the outermost component of the secretion machine, may be able to sense cell contact and somehow transduce a signal to the cytoplasmic side of the secretion machine to trigger its activity. Although there is no experimental support for this hypothesis, it is intriguing that discrete mutations have been found in the needle protein that result in constitutive type III secretion^{60,61}. It is therefore conceivable that these mutations lead to the assembly of needles whose conformation resembles the 'stimulated' state. Because monomers of the needle protein assemble into a helical structure, it has been proposed that sensing by this structure may occur through changes in the helical packing. However, mutants of the needle protein that showed altered secretion did not exhibit altered helical packing⁶². Whatever the sensing mechanism may be, the signal must be transduced to the cytoplasmic side of the secretion apparatus to 'open' the secretion machine. A candidate platform to transduce the activating signal within the bacterial cytoplasm is a multiprotein complex that in *Yersinia* spp. is composed of the YopN, SycN, YscB and TyeA proteins^{63–65}. This complex is essential to prevent the secretion of effector proteins in the absence of stimulatory signals. Homologues of components of this platform that perform similar functions have been identified in other T3SSs^{66,67}, indicating that this platform is presumably conserved in all T3SSs.

Once the effector proteins are released, evidence suggests that regulatory mechanisms are in place to ensure that the system is 'reset' and the secretion machine is 're-loaded' either with the same arsenal of effectors or, in some cases perhaps with new ones⁶⁸. Although the regulatory systems seem specific for each T3SS, a common mechanism involves the use of regulatory proteins that themselves are substrates of the T3SS (refs 69–71). Therefore, by removing either activators or repressors through secretion, T3SSs can couple the actual secretion process with the regulation of expression of effector proteins. A similar mechanism has been described in the flagellar system⁷².

Mimicry as a strategy to modulate cellular functions

It is clear that each T3SS delivers a unique arsenal of effector proteins, which have presumably been assembled and optimized through evolution to suit the specific needs of the bacteria that harbour them. Therefore, it is not surprising that proteins delivered by different T3SSs can modulate or interfere with a vast array of cellular functions including actin and tubulin dynamics, gene expression, vesicular trafficking, programmed cell death and cell cycle progression. The function of only a very small number of effector proteins has been characterized in some detail. Space limitations, however, prevent us from discussing the individual functions of these effector proteins in any detail. However, one general theme that emerges from the functional characterization and the atomic structures of a handful of these effector proteins is one of 'mimicry' as a central strategy to modulate cellular functions⁷³. Unlike other bacterial toxins that exert their function by introducing covalent, non-reversible modifications of their target host cell proteins, T3SS effectors seem to act by mimicking the function of host cell proteins. Indeed, this strategy seems appropriate to have been adapted by bacteria that have type III secretion systems as a central element for the establishment of a close functional interface that is often symbiotic in nature.

Another theme that has emerged from these studies is that these 'eukaryotic protein mimics' often seem to be the product of convergent evolution rather than the result of horizontal 'hijacking' of mammalian cell genes. For example, the *Salmonella* SPI-1 T3SS effector protein SopE is a Rho-family GTPase exchange factor

(GEF) that shares no sequence or structural similarity with eukaryotic GEFs⁷⁴. However, the crystal structure of the complex of SopE with its target Rac1 showed that the interaction leads to an outcome (that is, conformational changes in the critical switch 1 and switch 2 regions of Rac1) that is nearly indistinguishable from that of the interaction of a bona fide eukaryotic GEF and the same target⁷⁵. A similar example is that of a family of related GTPase activating proteins (GAPs) for Rho-family GTPases encoded by *Salmonella* spp., *Yersinia* spp. and *Pseudomonas aeruginosa* (SptP, YopE and ExoS, respectively)^{76–78}. Despite the utterly different amino acid sequence and structure of these GAPs in comparison with eukaryotic GAPs, they carry out their enzymatic activity using even the same chemistry as that used by their eukaryotic counterparts^{79,80}. A corollary of this observation is that it is very difficult to predict the actual function of most T3SS effector proteins from amino acid sequence analysis or even from the solution of their atomic proto-structures (that is, the atomic structures of the effectors in the absence of their host cellular targets). Indeed, most known or predicted T3SS effector proteins share no obvious amino acid sequence similarity to other proteins, except their putative orthologues and paralogues. Therefore, in most cases, understanding of the function of these effector proteins will require systematic functional analysis.

Yet another important theme that has emerged from the, so far, limited studies of T3SS effector proteins is that often the activities of different effectors delivered by the same machine are carefully coordinated and temporally regulated^{81,82}. Therefore, an understanding of the biology of a given T3SS will require an understanding of the biology of the function of all or most of the proteins that the system delivers into host cells. Indeed, understanding of the function of a given effector protein outside of the context of the function of the other effectors delivered by the same T3SS, may be incomplete, at best, or even entirely misleading. For example, the GAP activity of the *S. typhimurium* protein SptP by itself was originally interpreted as an activity aimed at disrupting the actin cytoskeleton of the target cell; however, in the context of its delivery along with activators of Rho-family GTPases, the function of SptP in *S. typhimurium* proved to be the preservation of the actin cytoskeleton rather than its disruption⁷⁶. Therefore, a comprehensive understanding of the function of most effectors delivered by a given T3SS may be required to ultimately know the actual function of that T3SS.

Future perspectives

The discovery of type III secretion machines has arguably been one of the most significant discoveries in bacterial pathogenesis of the past few years. The widespread distribution of such a macromolecular machine and its use in rather diverse biological contexts is a testament to the success of the evolutionary forces working to shape the complex functional interface between pathogenic or symbiotic bacteria and their eukaryotic hosts. Its central role in the interaction of many pathogenic bacteria opens up the possibility of developing new anti-infective strategies⁸³. In addition, a detailed understanding of these machines is allowing them to be harnessed to deliver heterologous proteins for therapeutic or vaccine purposes⁸⁴. The past few years have seen a rather remarkable increase in the understanding of these machines. There is no doubt that the importance and intrinsic beauty of these fascinating machines will continue to attract the attention of scientists and therefore progress is likely to continue at an even faster pace.

1. Pallen, M., Beatson, S. & Bailey, C. Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiol. Rev.* **29**, 201–229 (2005).
2. Saier, M. J. Evolution of bacterial type III protein secretion systems. *Trends Microbiol.* **12**, 113–115 (2004).
3. Macnab, R. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta* **1694**, 207–217 (2004).
4. Cornelis, G. The *Yersinia* Ysc–Yop 'type III' weaponry. *Nature Rev. Mol. Cell Biol.* **3**, 742–752 (2002).

5. Alfano, J. & Collmer, A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.* **42**, 385–414 (2004).
6. Kubori, T. *et al.* Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602–605 (1998).
7. Blocker, A. *et al.* Structure and composition of the *Shigella flexneri* “needle complex”, a part of its type III secretion. *Mol. Microbiol.* **39**, 652–663 (2001).
8. Sekiya, K. *et al.* Supramolecular structure of the enteropathogenic *Escherichia coli* type III secretion system and its direct interaction with the EspA-sheath-like structure. *Proc. Natl Acad. Sci. USA* **98**, 11638–11643 (2001).
9. Daniell, S. *et al.* The filamentous type III secretion translocon of enteropathogenic *Escherichia coli*. *Cell. Microbiol.* **3**, 865–871 (2001).
10. Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040–1042 (2004).
11. Sukhan, A., Kubori, T., Wilson, J. & Galán, J. E. Genetic analysis of assembly of the *Salmonella enterica* serovar Typhimurium type III secretion-associated needle complex. *J. Bacteriol.* **183**, 1159–1167 (2001).
12. Kubori, T., Sukhan, A., Aizawa, S. I. & Galán, J. E. Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. *Proc. Natl Acad. Sci. USA* **97**, 10225–10230 (2000).
13. Kimbrough, T. & Miller, S. Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation. *Proc. Natl Acad. Sci. USA* **97**, 11008–11013 (2000).
14. Crago, A. & Koronakis, V. *Salmonella* InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization. *Mol. Microbiol.* **30**, 47–56 (1998).
15. Daefler, S. & Russel, M. The *Salmonella typhimurium* InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Mol. Microbiol.* **28**, 1367–1380 (1998).
16. Crepin, V. *et al.* Structural and functional studies of the enteropathogenic *Escherichia coli* type III needle complex protein EscJ. *Mol. Microbiol.* **55**, 1658–1670 (2005).
17. Yip, C. *et al.* Structural characterization of the molecular platform for type III secretion system assembly. *Nature* **435**, 702–707 (2005).
18. Blocker, A., Komoriya, K. & Aizawa, S. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc. Natl Acad. Sci. USA* **100**, 3027–3030 (2003).
19. Morita-Ishihara, T. *et al.* Shigella Spa33 is an essential C-ring component of type III secretion machinery. *J. Biol. Chem.* **281**, 599–560 (2006).
20. Li, J. *et al.* Relationship between evolutionary rate and cellular location among the Inv/Spa invasion proteins of *Salmonella enterica*. *Proc. Natl Acad. Sci. USA* **92**, 7252–7256 (1995).
21. Collazo, C. & Galán, J. E. Requirement of exported proteins for secretion through the invasion-associated Type III system in *Salmonella typhimurium*. *Infect. Immun.* **64**, 3524–3531 (1996).
22. Journet, L., Agrain, C., Broz, P. & Cornelis, G. R. The needle length of bacterial injectisomes is determined by a molecular ruler. *Science* **302**, 1757–1760 (2003).
23. Hirano, T., Yamaguchi, S., Oosawa, K. & Aizawa, S. Roles of FlhK and FlhB in determination of flagellar hook length in *Salmonella typhimurium*. *J. Bacteriol.* **176**, 5439–5449 (1994).
24. Makishima, S., Komoriya, K., Yamaguchi, S. & Aizawa, S. Length of the flagellar hook and the capacity of the type III export apparatus. *Science* **291**, 2411–2413 (2001).
25. Crepin, V., Shaw, R., Abe, C., Knutton, S. & Frankel, G. Polarity of enteropathogenic *Escherichia coli* EspA filament assembly and protein secretion. *J. Bacteriol.* **187**, 2881–2889 (2005).
26. Marlovits, T. C. *et al.* Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**, 637–640 (2006).
27. Edqvist, P. *et al.* YscP and YscU regulate substrate specificity of the *Yersinia* type III secretion system. *J. Bacteriol.* **185**, 2259–2266 (2003).
28. West, N. *et al.* Optimization of virulence functions through glucosylation of *Shigella* LPS. *Science* **307**, 1313–1317 (2005).
29. Mota, L. J., Journet, L., Sorg, I., Agrain, C. & Cornelis, G. R. Bacterial injectisomes: needle length does matter. *Science* **307**, 1278 (2005).
30. Pettersson, J. *et al.* Modulation of virulence factor expression by pathogen target cell contact. *Science* **273**, 1231–1233 (1996).
31. Wulff-Strobel, C., Williams, A. & Straley, S. LcrQ and SycH function together at the Ysc type III secretion system in *Yersinia pestis* to impose a hierarchy of secretion. *Mol. Microbiol.* **43**, 411–423 (2002).
32. Sorg, J., Miller, N. & Schneewind, O. Substrate recognition of type III secretion machines—testing the RNA signal hypothesis. *Cell. Microbiol.* **7**, 1217–1225 (2005).
33. Sory, M.-P., Boland, A., Lambermont, I. & Cornelis, G. Identification of the YopE and YopH domains required for secretion and internalization into the cytosol of macrophages, using the *cyaA* gene fusion approach. *Proc. Natl Acad. Sci. USA* **92**, 11998–12002 (1995).
34. Schesser, K., Frithz-Lindsten, E. & Wolf-Watz, H. Delineation and mutational analysis of the *Yersinia pseudotuberculosis* YopE domains which mediate translocation across bacterial and eukaryotic cellular membranes. *J. Bacteriol.* **178**, 7227–7233 (1996).
35. Anderson, D. M. & Schneewind, O. A mRNA signal for the type III secretion of Yop proteins by *Yersinia enterocolitica*. *Science* **278**, 1140–1143 (1997).
36. Lloyd, S., Sjöström, M., Andersson, S. & Wolf-Watz, H. Molecular characterization of type III secretion signals via analysis of synthetic N-terminal amino acid sequences. *Mol. Microbiol.* **43**, 51–59 (2002).
37. Russmann, H., Kubori, T., Sauer, J. & Galán, J. Molecular and functional analysis of the type III secretion signal of the *Salmonella enterica* InvJ protein. *Mol. Microbiol.* **46**, 769–779 (2002).
38. Lee, S. H. & Galán, J. E. *Salmonella* type III secretion-associated chaperones confer secretion-pathway specificity. *Mol. Microbiol.* **51**, 483–495 (2004).
39. Wattiau, P. & Cornelis, G. R. SycE, a chaperone-like protein of *Yersinia enterocolitica* involved in the secretion of YopE. *Mol. Microbiol.* **8**, 123–131 (1993).
40. Feldman, M. & Cornelis, G. The multitasking type III chaperones: all you can do with 15 kDa. *FEMS Microbiol. Lett.* **219**, 151–158 (2003).
41. Stebbins, C. E. & Galán, J. E. Priming virulence factors for delivery into the host. *Nature Rev. Mol. Biol.* **4**, 738–743 (2003).
42. Woestyn, S., Sory, M. P., Boland, A., Lequenne, O. & Cornelis, G. R. The cytosolic SycE and SycH chaperones of *Yersinia* protect the region of YopE and YopH involved in translocation across eukaryotic cell membranes. *Mol. Microbiol.* **20**, 1261–1271 (1996).
43. Stebbins, C. E. & Galán, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414**, 77–81 (2001).
44. Birtalan, S. C., Phillips, R. M. & Ghosh, P. Three-dimensional secretion signals in chaperone-effector complexes of bacterial pathogens. *Mol. Cell* **9**, 971–980 (2002).
45. Lilic, M., Vujanac, M. & Stebbins, C. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Mol. Cell* **21**, 653–664 (2006).
46. Luo, Y. *et al.* Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nature Struct. Biol.* **8**, 1031–1036 (2001).
47. Gauthier, A. & Finlay, B. B. Translocated intimin receptor and its chaperone interact with ATPase of the type III secretion apparatus of enteropathogenic *Escherichia coli*. *J. Bacteriol.* **185**, 6747–6755 (2003).
48. Akeda, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, 911–915 (2005).
49. Muller, S. *et al.* Double hexameric ring assembly of the type III protein translocase ATPase HrcN. *Mol. Microbiol.* **61**, 119–125 (2006).
50. Sauer, R. *et al.* Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell* **119**, 9–18 (2004).
51. Hoiczky, E. & Blobel, G. Polymerization of a single protein of the pathogen *Yersinia enterocolitica* into needles punctures eukaryotic cells. *Proc. Natl Acad. Sci. USA* **98**, 4669–4674 (2001).
52. Sory, M.-P. & Cornelis, G. R. Translocation of a hybrid YopE-adenylate cyclase from *Yersinia enterocolitica* into HeLa cells. *Mol. Microbiol.* **14**, 583–594 (1994).
53. Hakansson, S. *et al.* The YopB protein of *Yersinia pseudotuberculosis* is essential for the translocation of Yop effector proteins across the target cell plasma membrane and displays a contact-dependent membrane disrupting activity. *EMBO J.* **15**, 5812–5823 (1996).
54. Tardy, F. *et al.* *Yersinia enterocolitica* type III secretion-translocation system: channel formation by secreted Yops. *EMBO J.* **18**, 6793–6799 (1999).
55. Mueller, C. *et al.* The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* **310**, 674–676 (2005).
56. Knutton, S. *et al.* A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J.* **17**, 2166–2176 (1998).
57. Roine, E. *et al.* Hrp pilus: an hrp-dependent bacterial surface appendage produced by *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl Acad. Sci. USA* **94**, 3459–3464 (1997).
58. Francis, M., Wolf-Watz, H. & Forsberg, A. Regulation of type III secretion systems. *Curr. Opin. Microbiol.* **5**, 166–172 (2002).
59. Ménard, R., Sansonetti, P. J. & Parsot, C. The secretion of the *Shigella flexneri* Ipa invasins is induced by the epithelial cell and controlled by IpaB and IpaD. *EMBO J.* **13**, 5293–5302 (1994).
60. Torruellas, J., Jackson, M., Pennock, J. & Plano, G. The *Yersinia pestis* type III secretion needle plays a role in the regulation of Yop secretion. *Mol. Microbiol.* **57**, 1719–1733 (2005).
61. Kenjale, R. *et al.* The needle component of the type III secretion of *Shigella* regulates the activity of the secretion apparatus. *J. Biol. Chem.* **280**, 42929–42937 (2005).
62. Cordes, F. *et al.* Helical packing of needles from functionally altered *Shigella* type III secretion systems. *J. Mol. Biol.* **354**, 206–211 (2005).
63. Iriarte, M. *et al.* TyeA, a protein involved in control of Yop release and in translocation of *Yersinia* Yop effectors. *EMBO J.* **17**, 1907–1918 (1998).
64. Day, J. & Plano, G. A complex composed of SycN and YscB functions as a specific chaperone for YopN in *Yersinia pestis*. *Mol. Microbiol.* **30**, 777–788 (1998).
65. Schubot, F. *et al.* Three-dimensional structure of a macromolecular assembly that regulates type III secretion in *Yersinia pestis*. *J. Mol. Biol.* **346**, 1147–1161 (2005).
66. Kubori, T. & Galán, J. E. *Salmonella* type III secretion-associated protein InvE controls translocation of effector proteins into host cells. *J. Bacteriol.* **184**, 4699–4708 (2002).
67. O’Connell, C. *et al.* SepL, a protein required for enteropathogenic *Escherichia coli* type III translocation, interacts with secretion component SepD. *Mol. Microbiol.* **52**, 1613–1625 (2004).

68. Demers, B., Sansonetti, P. J. & Parsot, C. Induction of type III secretion in *Shigella flexneri* is associated with differential control of transcription of genes encoding secreted proteins. *EMBO J.* **17**, 2894–2903 (1998).
69. Parsot, C. *et al.* A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*. *Mol. Microbiol.* **56**, 1627–1635 (2005).
70. Rietsch, A., Vallet-Gely, I., Dove, S. & Mekalanos, J. ExsE, a secreted regulator of type III secretion genes in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **102**, 8006–8011 (2005).
71. Urbanowski, M., Lykken, G. & Yahr, T. A secreted regulatory protein couples transcription to the secretory activity of the *Pseudomonas aeruginosa* type III secretion system. *Proc. Natl Acad. Sci. USA* **102**, 9930–9935 (2005).
72. Hughes, K. T., Gillen, K. L., Semon, M. J. & Karlinsey, J. E. Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. *Science* **262**, 1277–1280 (1993).
73. Stebbins, C. E. & Galán, J. E. Structural mimicry in bacterial virulence. *Nature* **412**, 701–705 (2001).
74. Hardt, W.-D., Chen, L.-M., Schuebel, K. E., Bustelo, X. R. & Galán, J. E. *Salmonella typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815–826 (1998).
75. Buchwald, G. *et al.* Structural basis for the reversible activation of a Rho protein by the bacterial toxin SopE. *EMBO J.* **21**, 3286–3295 (2002).
76. Fu, Y. & Galán, J. E. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
77. Von Pawel-Rammingen, U. *et al.* GAP activity of the *Yersinia* YopE cytotoxin specifically targets the Rho pathway: a mechanism for disruption of actin microfilament structure. *Mol. Microbiol.* **36**, 737–748 (2000).
78. Goehring, U., Schmidt, G., Pederson, K., Aktories, K. & Barbieri, J. The N-terminal domain of *Pseudomonas aeruginosa* exoenzyme S is a GTPase-activating protein for Rho GTPases. *J. Biol. Chem.* **274**, 36369–36372 (1999).
79. Stebbins, C. E. & Galán, J. E. Modulation of host signaling by a bacterial mimic: structure of the *Salmonella* effector SptP bound to Rac1. *Mol. Cell* **6**, 1449–1460 (2000).
80. Wurtele, M. *et al.* How the *Pseudomonas aeruginosa* ExoS toxin downregulates Rac. *Nature Struct. Biol.* **8**, 23–26 (2001).
81. Kubori, T. & Galán, J. E. Temporal regulation of *Salmonella* virulence effector function by proteasome-dependent protein degradation. *Cell* **115**, 333–342 (2003).
82. Kim, M. *et al.* Two *Pseudomonas syringae* type III effectors inhibit RIN4-regulated basal defense in *Arabidopsis*. *Cell* **121**, 749–759 (2005).
83. Nordfelth, R., Kauppi, A., Norberg, H., Wolf-Watz, H. & Elofsson, M. Small-molecule inhibitors specifically targeting type III secretion. *Infect. Immun.* **73**, 3104–3114 (2005).
84. Russmann, H. *et al.* Delivery of epitopes by the *Salmonella* type III secretion system for vaccine development. *Science* **281**, 565–568 (1998).

Acknowledgements We thank T. Marlovits and E. Stebbins for generously providing the images presented in Fig. 1 and 4, and members of the Galán laboratory for critical review of the manuscript. Work in the laboratory of J. G. is supported by the NIH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to J.E.G. (jorge.galan@yale.edu).

ARTICLES

Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs

Maurilio Sampaolesi^{1,2*}, Stephane Blot^{3*}, Giuseppe D'Antona², Nicolas Granger³, Rossana Tonlorenzi¹, Anna Innocenzi¹, Paolo Mognol⁴, Jean-Laurent Thibaud³, Beatriz G. Galvez¹, Ines Barthélémy³, Laura Perani¹, Sara Mantero⁴, Maria Guttinger⁵, Orietta Pansarasa², Chiara Rinaldi², M. Gabriella Cusella De Angelis², Yvan Torrente⁶, Claudio Bordignon¹, Roberto Bottinelli² & Giulio Cossu^{1,5,7}

Duchenne muscular dystrophy remains an untreatable genetic disease that severely limits motility and life expectancy in affected children. The only animal model specifically reproducing the alterations in the dystrophin gene and the full spectrum of human pathology is the golden retriever dog model. Affected animals present a single mutation in intron 6, resulting in complete absence of the dystrophin protein, and early and severe muscle degeneration with nearly complete loss of motility and walking ability. Death usually occurs at about 1 year of age as a result of failure of respiratory muscles. Here we report that intra-arterial delivery of wild-type canine mesoangioblasts (vessel-associated stem cells) results in an extensive recovery of dystrophin expression, normal muscle morphology and function (confirmed by measurement of contraction force on single fibres). The outcome is a remarkable clinical amelioration and preservation of active motility. These data qualify mesoangioblasts as candidates for future stem cell therapy for Duchenne patients.

Duchenne muscular dystrophy primarily affects skeletal muscle, causing fibre degeneration, progressive paralysis and death¹. No effective treatment exists although novel therapeutic strategies, ranging from new drugs to gene and cell therapy, hold promise for significant advance in the future². In particular, different types of stem cell have been shown to induce dystrophin synthesis and partial rescue of the pathology in dystrophic mice^{3–8}. However, dystrophic mice do not display clinical signs of the disease, and to proceed to a clinical trial it is imperative to show efficacy in a large, non-syngeneic animal model of muscular dystrophy. Golden retriever muscular dystrophy (GRMD)^{9,10} is a very severe form of dystrophy, which

affects not only limb, respiratory and heart muscles but also pharyngeal muscles, resulting in a severe involvement of the digestive tract; although variability exists between individuals, by 8 months of age most dogs walk with great difficulty (Supplementary Movie 1). To test the efficacy of cell or gene therapy, we transplanted GRMD dogs with either autologous genetically corrected or donor wild-type mesoangioblasts, under different regimes of immune suppression.

Ten dystrophic dogs were treated in three experiments and a general scheme of treatments and outcome is reported in Table 1. Four dogs received autologous mesoangioblasts, transduced *in vitro* with a lentiviral vector expressing human microdystrophin (Supplementary

Table 1 | Summary of treatment

Dog no.	Dog name	Cell treatment	Lentiviral vector	Onset of treatment	Immune suppression (time)	Dystrophin expression	Motility	Outcome of experiment (at time P400)
01A	Ucal	Autologous, gene therapy	CK- μ dys-ires-GFP	P118	–	+ / –	Loss	Euthanasia (P272)
02H	Vrillie	Heterologous, WT donor	–	P80	CYC A (P78)	+	Loss	Euthanasia (P235)
03H	Valgus	Heterologous, WT donor	–	P75	CYC A (P73)	+++	No decline	Alive and well
04H	Varus	Heterologous, WT donor	–	P75	RAP (P73)	+++	Modest decline	Alive and well
05H	Viko	Heterologous, WT donor	–	P77	RAP + IL-10 (P74)	ND	ND (sudden death)	Myocarditis (P186)
06A	Vaccin	Autologous, gene therapy	MLC1F- μ dys	P113	–	++	Major decline	Euthanasia (P326)
07A	Valium	Autologous, gene therapy	MLC1F- μ dys	P113	–	ND	Loss	Pneumonia (P245)
08A	Vampire	Autologous, gene therapy	MLC1F- μ dys	P113	–	++	Major decline	Pneumonia (P154)
09H	Azur	Heterologous, WT donor	–	P159	CYC A (P157)	++	Restored	Alive and well
10H	Azor	Heterologous, WT donor	–	P159	CYC A (P157)	+++	Restored	Alive and well
11U	Akan	None	–	–	–	–	Loss	Euthanasia (P380)
12U	Vulcano	None	–	–	–	–	Loss	Euthanasia (P376)
13U	Viking	None	–	–	–	–	Loss	Euthanasia (P340)

Each dog was given a specific name and a sequential number, followed by A (transplanted with autologous cells), H (transplanted with heterologous cells) or U (untreated). The nature of the lentiviral vector is indicated (CK- μ dys-ires-GFP, creatine kinase promoter driving microdystrophin-ires-GFP; MLC1F- μ dys, myosin light chain 1 fast promoter driving microdystrophin). Dystrophin expression was quantified as follows: –, average dystrophin (or micro-dystrophin) expression in less than 1% of positive fibres; + / –, less than 10% of positive fibres; +, less than 20% of positive fibres; ++, less than 50% of positive fibres; + + +, more than 50% of positive fibres. CYC A, cyclosporine; ND, not determined; RAP, rapamycin; WT, wild-type. Euthanasia was administered when clinical conditions worsened.

¹San Raffaele Scientific Institute, Università Vita e Salute, Stem Cell Research Institute, Via Olgettina 58, 20132 Milan, Italy. ²Department of Experimental Medicine and Interuniversity Institute of Myology, University of Pavia, Via Forlanini 6-8, 27100 Pavia, Italy. ³Neurobiology Laboratory, École Vétérinaire d'Alfort, 7 Avenue Général de Gaulle, 94704 Maisons-Alfort cedex, France. ⁴Department of Bioengineering, Politecnico di Milano, Piazza Leonardo Da Vinci, 20130 Milan, Italy. ⁵Institute of Cell Biology and Tissue Engineering, San Raffaele Biomedical Science Park of Rome, Via Castel Romano 100, 00128 Rome, Italy. ⁶IRCCS Fondazione Policlinico di Milano, Department of Neurological Sciences, University of Milan, Via Sforza 35, 20122 Milan, Italy. ⁷Department of Biology and Centre for Stem Cell Research, University of Milan, Via Celoria 28, 20130 Milan, Italy.

*These authors contributed equally to this work.

Fig. 1), a truncated but functional version of dystrophin¹¹; six dogs received wild-type mesoangioblasts from a single DLA (dog leukocyte antigen)-unrelated donor under treatment with either cyclosporine or rapamycin. Three other dystrophic dogs were not treated and served as controls. All dystrophic dogs received steroids daily as standard treatment. Overall, the results showed that donor wild-type mesoangioblasts significantly ameliorate many symptoms of canine muscular dystrophy, whereas autologous genetically corrected cells are much less effective.

Isolation and characterization of canine mesoangioblasts

Both wild-type and dystrophic mesoangioblasts were isolated from the outgrowth of small, vessel-containing, tissue fragments from muscle biopsies performed after diagnosis, at about 15 days postnatal (P15). The cells show a morphology very similar to that of mouse mesoangioblasts¹² (Fig. 1a), proliferate efficiently in a medium devised for stem cells (Fig. 1b) and show a euploid karyotype of 78 chromosomes both at early and late passage (Fig. 1c); cells undergo senescence after about 25 population doublings. Canine mesoangioblasts express CD44 and CD13 but not CD34, CD45, CD117 or CD31 (not shown). Dog mesoangioblasts differentiate into multinucleated myotubes when co-cultured with C2C12 mouse myoblasts or when transfected with MyoD. For the gene transfer experiments, proliferating mesoangioblasts isolated from dystrophic dogs were transduced with lentiviral vectors expressing human microdystrophin and (only for Ucal, the first dog treated) enhanced green fluorescent protein (EGFP). Both proteins became readily detectable after myogenic differentiation (Fig. 1d–f). Finally, to test the ability of these cells to reconstitute muscle fibres *in vivo*, both wild-type and GRMD genetically corrected mesoangioblasts were injected into SCID (severe combined immunodeficiency)-mdx mice, which do not reject xenogenic cells; the cells migrated from the femoral artery to the downstream muscles with an efficiency similar to that of their wild-type mouse counterparts¹³ (Fig. 1g). Three weeks after injection, canine mesoangioblasts gave rise to dystrophin-positive fibres containing dog nuclei, identified by anti-human lamin A-C antibody, which recognizes human and dog but not mouse nuclei (Fig. 1h, i). Thus, dog mesoangioblasts seem similar to their mouse postnatal counterparts by all the parameters tested, with the notable exception of a finite lifespan, a predictable difference between cells from rodents and other mammals.

Feasibility experiment

Dogs are identified by name and also by a sequential number followed by a letter (A for autologous cell transplantation, H for heterologous cell transplantation, and U for untreated) (Table 1).

Two dogs, Ucal and Vrillie, were treated with three consecutive (at 1-month intervals) injections of 5×10^7 cells into the femoral artery. Ucal (01A) received autologous cells, transduced with the lentiviral vector expressing human microdystrophin (Supplementary Fig. 1a). Vrillie (02H) received wild-type donor cells under a regimen of cyclosporine treatment.

During and after the treatment, Ucal (01A) and Vrillie (02H) did not show appreciable sign of clinical amelioration and underwent a progressive decline in their walking ability. Biopsies, taken 1 month after the third injection, revealed variable morphology in different muscles of the injected legs, varying from severe and advanced degeneration in the tibialis cranialis of Ucal (01A), shown in Fig. 2a, to an intermediate severity in the same muscle of Vrillie (02H), shown in Fig. 2b. In general, the morphology of Vrillie (02H) was better but still variable, with several areas being quite well preserved. Dystrophin expression also showed variability and was in general correlated with morphology. At 8 months of age, the proportion of revertant fibres in dystrophic dogs varies from 0.02% to 0.3% (ref. 14). In the biopsies collected, dystrophin-positive fibres ranged from 2% to 7% in Ucal (01A) and from 4% to 10% in Vrillie (02H) (not shown). Figure 3a, a' shows a biopsy of Ucal in which clusters of

dystrophin-positive fibres with several centrally located nuclei can be observed. Biopsies from contralateral, non-injected legs showed poorer morphology and a smaller proportion (2% or less) of dystrophin fibres. Unexpectedly, we found areas of dystrophin expression in the triceps brachialis of Vrillie (02H) (not shown), indicating that

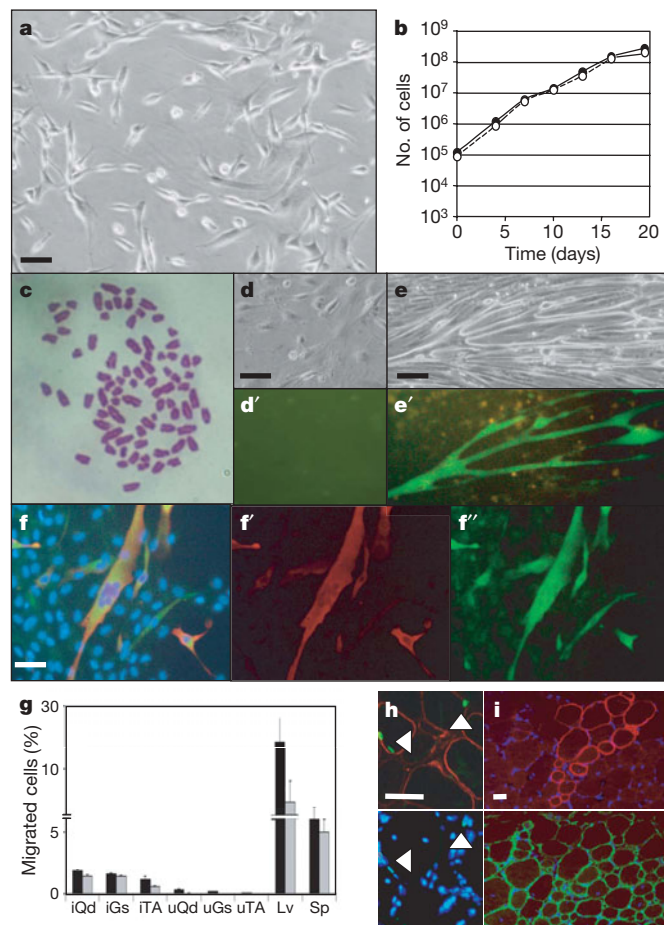


Figure 1 | Characterization of dog mesoangioblasts. **a**, Morphology of canine mesoangioblasts isolated from muscle biopsies of a golden retriever dog at P15. **b**, Proliferation curves of wild-type (filled circles) and dystrophic (open circles) canine mesoangioblasts. **c**, Karyotype of canine mesoangioblasts, which are consistently euploid until senescence. **d–f**, Transduction of dystrophic canine mesoangioblasts with a lentiviral vector expressing human microdystrophin and the EGFP gene reporter under the control of muscle-specific creatine kinase promoter. GFP expression is undetectable in proliferating mesoangioblasts (**d**, **d'**) but is readily detected in myotubes derived from the fusion of transduced mesoangioblasts with C2C12 myoblasts (**e**, **e'**). Similar results were obtained after MyoD-induced differentiation: GFP-positive cells (**f'**) also express MyHC (**f'**) as confirmed in the merged image (**f**), which also shows multinucleation. **g**, Migration of canine mesoangioblasts into skeletal muscle: 5×10^5 mouse (black bars) or dog (grey bars) mesoangioblasts, previously transduced with a GFP-expressing lentiviral vector⁴, were injected into the right femoral artery of SCID-mdx mice. Six hours after injection, several muscles were isolated and the presence of donor cells was measured by real-time PCR analysis for GFP as detailed elsewhere¹³. Qd, quadriceps; Gs, gastrocnemius; Tc, tibialis cranialis; Lv, liver; Sp, spleen; the letter i (injected) before the muscle name indicates muscle isolated from the injected leg; the letter u (uninjected) indicates muscles isolated from the contralateral leg. **h**, Top: immunofluorescence with antibodies against human lamin A-C (green) and dystrophin (red), revealing dog mesoangioblasts inside the muscle fibres of SCID-mdx mice, 21 days after intra-arterial injection. Bottom: nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI). **i**, Immunofluorescence of fibres with antibodies against dystrophin (red, top) and laminin (green, bottom) in the muscle of SCID-mdx mice, 21 days after intra-arterial injection. Scale bar, 20 μ m. Error bars, 1 s.d.

injected cells must to a certain extent transit through the capillaries of the injected leg and then through the filter organs, finally entering the arterial circulation to reach the forelimb muscles. Overall, the intensity of staining was weak; indeed, western blot analysis confirmed very low, barely detectable, levels of microdystrophin and dystrophin, respectively. Immunohistochemistry detected several inflammatory infiltrates containing mainly macrophages and lymphocytes (Supplementary Fig. 2a, b). Although these data demonstrate donor-cell-dependent dystrophin expression in dystrophic dogs and thus the feasibility of this therapeutic approach, the results were modest overall. To improve the efficacy of the treatment, the number of injections was increased to five (always at 1-month intervals), a new lentivector was produced in which muscle-specific creatine kinase was replaced by the stronger myosin light chain 1F promoter and GFP was deleted because we could not detect GFP in unfixed cryostat sections (Supplementary Fig. 1b).

Efficacy of heterologous wild-type mesoangioblasts

Three dogs, namely Valgus, Varus and Viko (03H, 04H and 05H, respectively), were treated with donor cells. One of these dogs (Valgus, 03H) received five arterial systemic injections (5×10^7 cells each) through a catheter that was introduced in the left femoralis artery and reached the aortic arch at the level of the left subclavia: cells were released mainly in the two large arteries. From a clinical point of view, Valgus (03H) had optimal performance and was still walking well 5 months after the last injection and the termination of immune suppression, at the age of 13 months (Supplementary Movie 2). Valgus (03H) was treated with cyclosporine, whereas the other two dogs, Varus (04H) and Viko (05H), were treated with rapamycin and with rapamycin and interleukin (IL)-10, respectively. Different protocols of immune suppression were tested to evaluate efficacy versus toxicity in this model, but the results did not show significant differences between cyclosporine and rapamycin. In fact, Varus (04H; Supplementary Movie 3) and Viko (05H) also had good clinical performance but after 2 months of treatment Viko (05H) died suddenly of a fulminans myocarditis whose cause remained unexplained. Varus (04H), in contrast, progressively lost walking ability after the

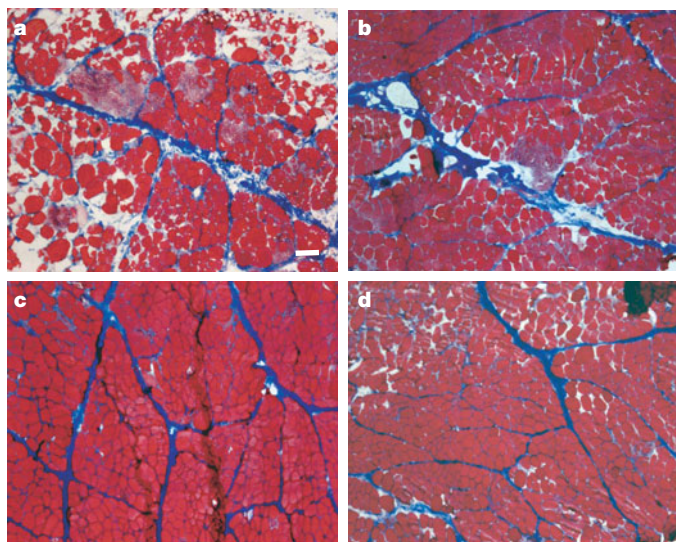


Figure 2 | Morphology of muscle in treated dogs. Azan Mallory staining of several muscle biopsies from the tibialis cranialis of dogs transplanted with heterologous mesoangioblasts or autologous, genetically corrected, mesoangioblasts. **a–c**, Examples show variability from severely affected tissue of Ucal (01A; **a**) and Vrilie (02H; **b**), with many infiltrates, collagen and fat deposition, to the almost normal appearance of tissue from Valgus (03H; **c**), showing only thickened interstitial tissue. **d**, A biopsy from Vaccin (06A) shows an intermediate situation with well-preserved morphology. Bar = 100 μ m.

end of immune suppression. Unexpectedly, however, this animal had no detectable anti-dystrophin antibodies and his circulating lymphocytes did not react to donor mesoangioblasts or to protein extracts from the transplanted muscle (Supplementary Fig. 1c, d).

At the end of the treatment, biopsies were taken from several muscles of the injected and contralateral leg of these dogs, treated with heterologous wild-type cells. Histological analysis of biopsies from Valgus (03H) revealed generally well-preserved morphology (Fig. 2c), although areas of degeneration and regeneration were detected infrequently. Supplementary Fig. 3 shows a large area of the tibialis cranialis of Valgus (03H) (Supplementary Fig. 3a, a'), better preserved than a corresponding area from Vampire (08A), a dog transplanted with autologous cells (Supplementary Fig. 3b, b'),

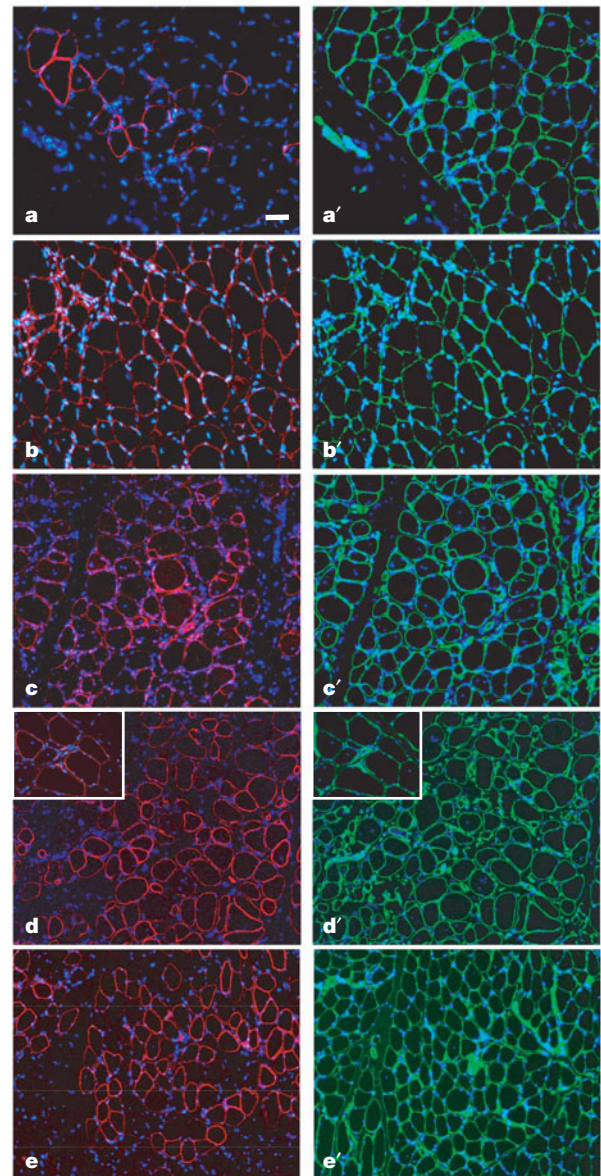


Figure 3 | Immunofluorescence analysis of tissue from treated dogs. Double immunofluorescence analysis of muscle biopsies from the tibialis cranialis of Ucal (01A; **a**, **a'**), Valgus (03H; **b**, **b'**), Varus (04H; **c**, **c'**), Vampire (08A; **d**, **d'**) and Azor (10H; **e**, **e'**), stained with anti-laminin antibody (green in **a'–e'** and inset in **d'**) and anti-dystrophin (red in **a–d**) or anti- β -sarcoglycan (red in inset in **d**). Nuclear staining with DAPI appears in blue. Note clusters of dystrophin-expressing fibres in Ucal, 01A (**a**) and extensive reconstitution of dystrophin-expressing fibres in Valgus (03H; **b**), Varus (04H; **c**) and Azor (10H; **e**). Several fibres expressing microdystrophin in Vampire, 08H (**d**) also express β -sarcoglycan in the same fibres stained on serial, non-adjacent sections (inset in **d**). Scale bar, 100 μ m.

described in detail below, which is itself better preserved than the biopsy from Akan (11U), an untreated dog (Supplementary Fig. 3c, c'). Both forelimb muscles (not shown) and the diaphragm (Supplementary Fig. 5a) of Valgus (03H) also showed a relatively preserved morphology. Similarly, dystrophin was expressed in most of the examined biopsies from several different muscles of Valgus (03H9), such as tibialis cranialis (Fig. 3b, b'), gastrocnemius and biceps femoralis (not shown). Large areas, containing several hundred dystrophin-expressing fibres (up to 70% of total fibres), were frequently detected: an example of the sartorius of Valgus (03H) is shown in Supplementary Fig. 5. Several clusters of dystrophin positive fibres could be detected even in the diaphragm (Supplementary Fig. 4e).

The biopsies of Varus (04H) also showed well-preserved morphology although signs of muscle necrosis and regeneration were still present (Fig. 2d). Overall, most of the biopsies in all muscles were morphologically less affected than those of untreated dogs. Dystrophin expression in these biopsies was variable but large areas of dystrophin-positive fibres could be detected, as shown in Fig. 3c, c'. In addition, the diaphragm of Varus was partly preserved (Supplementary Fig. 4b) and contained clusters of dystrophin-positive fibres (Supplementary Fig. 4f). Figure 4 shows the percentage of dystrophin-expressing fibres, which ranged from 10% to at least 70% in two distant sections of three different biopsies each of selected muscles from Valgus (03H) and Varus (04H). Western blot analyses of extracts from different biopsies of the same muscles confirmed the presence of significant amounts of dystrophin, varying from an undetectable signal to about 60% of a control wild-type canine muscle (Fig. 4c). When dystrophin was clearly detected, β -sarcoglycan was also detected, indicating reconstitution of the dystrophin-associated complex.

Efficacy of autologous mesoangioblasts

Three dogs, Vaccin, Valium and Vampire (06A, 07A and 08A, respectively), received their own mesoangioblasts, transduced with the new lentiviral vector expressing human microdystrophin. Like Ucal (01A) in the first experiment, they received the first injection at about 4 months of age (P113). Valium (07A) and Vampire (08A) died of pneumonia after the third and fifth injection, respectively. Because of early death, no force measurement was performed but a movie showing reduced walking ability in Vampire (08A) was taken few days before his death (Supplementary Movie 4) and biopsies were taken at autopsy. Valium (07A) received systemic arterial delivery, like Valgus (03H), and maintained a certain walking ability, albeit very stiff. The biopsies of the three dogs revealed partly preserved morphology; for example, tibialis cranialis from Vaccin (06A) is shown in Fig. 2d. In addition, microdystrophin expression in the tibialis cranialis of Vampire (08A) was significantly widespread (Fig. 3d, d'), despite the poor clinical performance. The microdystrophin was able to recruit at least some member of the dystrophin-associated complex such as β -sarcoglycan (shown in a serial, non-adjacent section in insets to Fig. 3d, d' and by western blot analysis in Fig. 4b). A quantitative analysis of microdystrophin for Vaccin (06A) and Vampire (08A) is reported in Fig. 4. Thus, all three dogs treated with autologous, genetically corrected cells performed poorly, even though two of them showed amelioration of morphology and microdystrophin-expressing fibres, ranging from 5% to 50%.

Efficacy of late transplantation of donor mesoangioblasts

To verify whether the less effective results obtained with autologous cells were due to the later onset of the treatment, two dogs, Azur (09H) and Azor (10H), received the first of five injections at P159 (5 months), 1 month later than those animals treated with their own genetically modified cells. Although both dogs were already severely affected at the onset of treatment, Azor (10H) showed a striking improvement of motility (while still limping) and was even able to

run at the end of treatment (Supplementary Movie 5). Azur (09H) showed a less evident but clearly detectable amelioration (Supplementary Movie 6). In particular, although there was variation between dogs in the progression of the disease, spontaneous improvement was never observed. Biopsies taken at the end of treatment also showed relatively well-preserved morphology (although with sclerosis and infiltrates) and widespread dystrophin expression in the tibialis cranialis (Fig. 4e, e'), biceps femoralis, gastrocnemius, sartorius and even diaphragm (Supplementary Fig. 6 and data not shown). Thus, we conclude that even with a later onset of treatment, donor cells seem to produce a greater amelioration of muscular dystrophy than is produced by autologous microdystrophin-expressing cells. After the end of treatment with cyclosporine, Azor (10H) continued to walk actively until the end of the experiment, whereas Azur (09H) rapidly lost walking ability, much as had previously been observed with Valgus (03H) and Varus (04H).

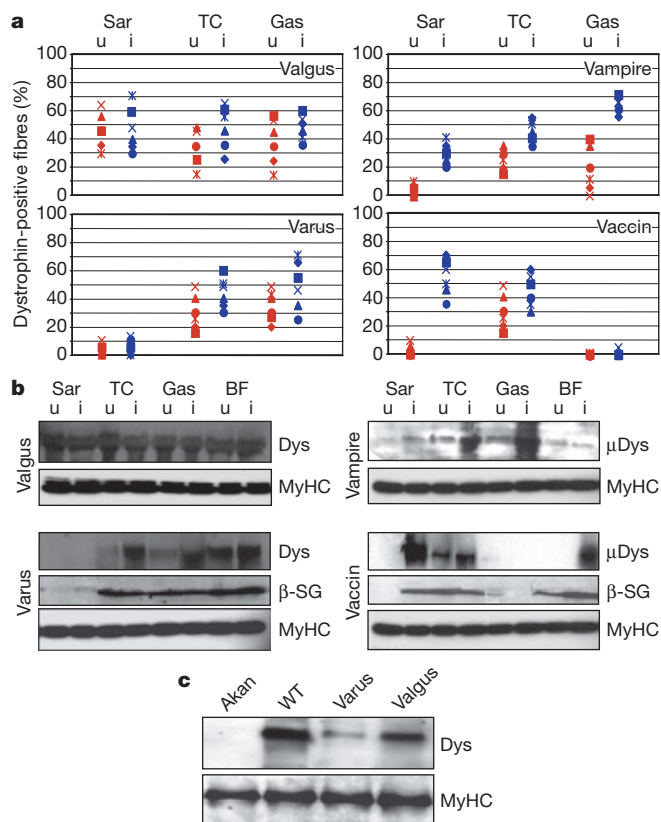


Figure 4 | Quantitative analysis of dystrophin content in tissue from treated dogs. **a**, Three separate sections for two different biopsies of the injected (i, blue symbols) and non-injected (u, red symbols) sartorius (Sar), tibialis cranialis (TC) and gastrocnemius (Gas) for two dogs injected with donor cells (Valgus (03H) and Varus (04H)) and two dogs injected with autologous, genetically corrected, cells (Vaccin (06A) and Vampire (08A)) were analysed for dystrophin expression. A total of 200 laminin-positive fibres (of any size) were counted in randomly selected fields, and the percentage of these fibres also expressing dystrophin was calculated. **b**, Western blot analysis of different biopsies from the muscles shown in **a** plus the biceps femoralis (BF) showing the expression of dystrophin (Dys, Valgus (03H) and Varus (04H)) or microdystrophin (μ Dys, Vaccin (06A) and Vampire (08A)). Each lane was loaded with 80 μ g of proteins. Myosin heavy chains (MyHC) are also shown as an internal standard in the bottom panel. For Varus (03H) and Vampire (08A) the expression of β -sarcoglycan (β -SG) is also shown in the middle panel. There is a good correlation between the expression of dystrophin or microdystrophin and β -sarcoglycan. **c**, Expression of dystrophin and myosin heavy chains in the injected biceps femoralis of GMRD untreated (Akan (U11)), wild type, Varus (04H) and Valgus (03H); only 30 μ g of total proteins was loaded in each lane for direct comparison.

Analysis of muscle-released serum enzymes

Throughout treatment, serum collected from treated dogs was used to detect the levels of creatine kinase and other enzymes released by damaged muscle fibres (Supplementary Fig. 7); in general, the injection of mesoangioblasts corresponded to rapid and often profound decrease in enzyme activity. This may be due to improved survival of fibres reconstituted with transplanted cells but also to factors such as insulin-like growth factor 1, basic fibroblast growth factor and others released by donor cells¹⁵ and contributing to fibre survival. At the end of treatment, the level of creatine kinase in all treated dogs was well below the those reported for this form of muscular dystrophy⁹. Other enzymes released by damaged muscle fibres, such as aspartate aminotransferase, showed an overall similar pattern (data not shown).

Enhancement of contraction force

To test whether morphological and biochemical changes would correspond to increased force of contraction we performed two different types of analysis: tetanic force of the tibialis cranialis and extensor digitorum longus muscle for Valgus (03H), Varus (04H),

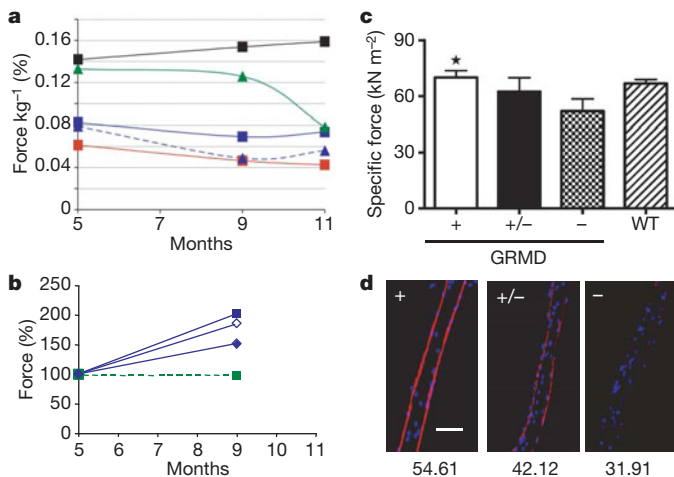


Figure 5 | Physiology. Functional properties of skeletal muscles *in vivo* and of individual muscle fibres *in vitro* of treated dystrophic dogs after five consecutive injections of donor wild-type mesoangioblasts. **a**, Tetanic force of the flexor muscles of the cranial tibial compartment of different dogs measured after maximal stimulation of the common peroneal nerve and recorded with an ergometer developed in house. Black squares, normal wild-type dog; red squares, dystrophic untreated dog (Akan 11U); green triangles, dystrophic dog transplanted with autologous genetically corrected mesoangioblasts (Vaccin 06A); blue symbols, dystrophic dogs transplanted with donor wild-type mesoangioblasts (squares, Valgus (03H) and Varus (04H); triangles, Azur (09H) and Azor (10H)). Tetanic force was normalized to each dog's body weight to obtain weight-corrected specific force. **b**, Relative increase in force of the flexor muscles of the cranial tibial compartment of the injected leg versus the untreated leg. Blue squares, Varus (04H); green squares, Vaccin (06A); filled blue diamonds, Azur (09H); open blue diamonds, Azor (10H). **c**, Specific force (maximum isometric force/cross-sectional area) of a population of 199 single muscle fibres dissected from tibialis cranialis, gastrocnemius, and sartorius muscles of Valgus (03H) and Varus (04H), indicated together as GRMD, and of 148 fibres dissected from the same muscles of two control golden retriever dogs (WT). All fibres shown are fast fibres (type 2A, 2AX or 2X) ($n = 71$ for GRMD; $n = 106$ for WT). Type 2A, 2AX and 2X fibres were pooled together because no difference in specific force was observed between these fast fibre types in either GRMD or WT dogs. Fibres from Valgus (03H) and Varus (04H) were grouped into dystrophin-positive (+), partly positive (+/-) and negative (-) on the basis of the presence of dystrophin detected, after force determination, by immunostaining by anti-dystrophin antibody. **d**, Examples of positive (+), partly positive (+/-) and negative (-) single skinned fibres immunostained for dystrophin after force determination. The specific force (maximum isometric force/cross sectional area) developed by each fibre is shown under each panel. Scale bar, 100 μ m. Error bars, 1 s.d.

Vaccin (06A), Azur (09H) and Azor (10H), and force of contraction on isolated single fibres from Varus (03H), Valgus (04H), Vaccin (06A) and Valium (07A). Figure 5a shows the tetanic force of a normal dog (black squares), which increases progressively with age. Dystrophic dogs have variable tetanic force (ranging from 90% to 40% of wild type) but, in the absence of treatment (red squares), progressively lose it. In the experiment reported, Vaccin (06A), who received autologous cells (green triangles), showed a strong tetanic force at the onset of treatment and maintained it up to the fourth injection but later showed a significant decrease. In contrast, the four dogs receiving donor cells (blue squares and triangles) started from a decreased tetanic force but maintained it though the treatment and eventually showed a modest increase in the last measurement, at the end of treatment. To avoid the problems deriving from this variability, force was also measured and reported as the percentage increase in treated over untreated legs. Figure 5b shows that all dogs receiving donor cells showed an increase in tetanic force ranging from 50% to 100% in the treated leg, whereas the dog receiving his own cells (Vaccin, 06A) did not show any increase. To measure force of contraction at the single-fibre level, a large population of single skinned fibres was dissected from the tibialis cranialis, gastrocnemius and sartorius muscles of Valgus (03H) and Varus (04H) and of two control wild-type dogs. After determination of specific force, each fibre was also analysed for the expression of dystrophin and of fast or slow myosin heavy chains (MyHC). The results reported in Fig. 5c show that fast fibres from dystrophic dogs had a complete recovery in force when expressing dystrophin, up to the level of fibres from wild-type dogs. Fibres showing partial expression of dystrophin also showed partial force recovery. A picture showing fully or partly reconstituted as well as dystrophin-negative fibres is shown in Fig. 5d. In contrast, dystrophic slow fibres did not show reduced force in comparison with wild-type fibres, as shown previously¹⁶ (data not shown). A similar measurement on the single skinned fibres of the tibialis cranialis and sartorius of Vaccin and Valium (06A and 07A) showed a similar trend, but the increase in force of contraction was modest and not statistically significant (data not shown).

Immune reaction against dystrophin and donor cells

To test the possible occurrence of an immune reaction in the transplanted dogs against donor cells and/or the transgene (dog dystrophin and human microdystrophin, respectively), we performed an immunocytochemical analysis of cellular infiltrates, a western blot analysis to test the reactivity of dog sera and a lymphocyte proliferation assay to test the appearance of cellular immunity against donor cells and/or the transplanted muscle tissue¹⁷⁻¹⁹. Results indicate a low frequency of infiltrates, an absence of serum antibodies and modest activation only for local lymphocytes exposed to transplanted muscle tissue (Supplementary Information).

Conclusions and future perspectives

Gene or cell therapy approaches for GRMD have until now produced negative²⁰ or modestly positive results²¹⁻²³. We show here that it is possible to transplant mesoangioblasts into dystrophic dogs and obtain an extensive reconstitution of fibres expressing dystrophin, an improvement in the contraction force and, in many cases, a preservation of walking ability. Previous work in the mouse^{4,13} showed that mesoangioblasts express some of the protein that leukocytes use to cross the vessel wall and so they invade the interstitial tissue, eventually to fuse with and contribute sarcoglycan to regenerating fibres. Donor wild-type mesoangioblasts seemed to be more efficient than autologous, genetically corrected cells. Possibly microdystrophin produces a modest functional rescue when delivered late through donor cells in contrast with the excellent functional rescue induced when delivered as a transgene¹¹. A different onset of treatment should not be crucial because two late-transplanted dogs (Azor

(09H) and Azur (10H)) showed improved motility; in addition, the differential treatment with cyclosporine should not matter because the drug itself had no effect on dogs transplanted with wild-type bone marrow¹⁹ (S.B. and C. Esciou, unpublished observation).

As mentioned above, extensive variability exists between dystrophic dogs; this is the situation that will be faced with patients, who also show phenotypic variability within the same mutation²⁴. Even taking this into account, four of six dogs treated with donor cells showed clear clinical amelioration; of the remaining two, Viko (05H) died early and could not be analysed in detail and Vrillie (02H) probably received too few cells. At the end of immune suppression, two of the four dogs showing amelioration, namely Valgus (03H) and Azor (10H), continued to walk well until the end of the experiment, whereas the other two, Varus (04H) and Azur (09H), rapidly lost walking ability. We do not have an explanation for this difference, which may reflect the different survival of transplanted organs after the end of immune suppression²⁵.

Extrapolation of these results to a possible future clinical trial would suggest starting with donor cells from an HLA-matched donor under a regime of immune suppression. The absence of cells of the immune system in the donor population should create a more favourable situation than with bone marrow transplantation. In fact, the modest immune reaction detected against donor canine mesoangioblasts, which were not DLA matched, indicates that immune suppression might be modest and perhaps transient.

Thus, the work reported here sets the logical premise for the start of clinical experimentation that may lead to an efficacious therapy for Duchenne muscular dystrophy.

METHODS

Cells and culture conditions. Dog mesoangioblasts were isolated from muscle biopsies of golden retriever dystrophic dogs (GRMD) and wild-type golden retrievers as described previously⁴. Details are given in Supplementary Information.

In vivo cell migration assay. Dog mesoangioblasts were analysed for the ability to migrate *in vivo* as described¹³. Details are given in Supplementary Information.

Fluorescence-activated cell sorting. For fluorescence-activated cell sorting, cells were detached and stained sequentially with primary antibodies against CD13, CD31, CD34, CD44, CD 117 and CD45 (Becton Dickinson) and immunofluorescent secondary antibodies and fixed with 2% paraformaldehyde until analysis with a FACSCalibur (Becton Dickinson).

Intra-artery delivery of mesoangioblasts in GRMD. About 5×10^7 mesoangioblasts were injected through the femoral artery of ten dogs, as detailed in Supplementary Information.

Biochemical and histological analysis. Immunoblotting, immunofluorescence and histological staining were performed as previously described⁴ and detailed in Supplementary Information.

In vivo analysis of muscle force. Maximal isometric force generated by flexor muscles of the tibialis cranialis compartment (tibialis cranialis and extensor digitorum longus) was measured by supramaximal percutaneous stimulation of the common peroneal nerve with a set-up detailed in Supplementary Information.

Clinical GRMD follow-up. Male golden retrievers carrying the GRMD mutation obtained from a colony at the veterinary school of Alfort were observed for disease for at least 1 month before being entered into the study. Details are given in Supplementary Information.

Received 4 April; accepted 26 September 2006.

Published online 15 November 2006.

- Emery, A. E. The muscular dystrophies. *Lancet* **359**, 687–695 (2002).
- Cossu, G. & Sampaioles, M. New therapies for muscular dystrophy: cautious optimism. *Trends Mol. Med.* **10**, 516–520 (2004).
- Qu-Petersen, Z. *et al.* Identification of a novel population of muscle stem cells in mice: potential for muscle regeneration. *J. Cell Biol.* **157**, 851–864 (2002).
- Sampaioles, M. *et al.* Cell therapy of α -sarcoglycan null dystrophic mice through intra-arterial delivery of mesoangioblasts. *Science* **301**, 487–492 (2003).
- Bachrach, E. *et al.* Systemic delivery of human microdystrophin to regenerating mouse dystrophic muscle by muscle progenitor cells. *Proc. Natl Acad. Sci. USA* **101**, 3581–3586 (2004).

- Torrente, Y. *et al.* Human circulating AC133⁺ stem cells restore dystrophin expression and ameliorate function in dystrophic skeletal muscle. *J. Clin. Invest.* **114**, 182–195 (2004).
- Rodriguez, A. M. *et al.* Transplantation of a multipotent cell population from human adipose tissue induces dystrophin expression in the immunocompetent mdx mouse. *J. Exp. Med.* **201**, 1397–1405 (2005).
- Dezawa, M. *et al.* Bone marrow stromal cells generate muscle cells and repair muscle degeneration. *Science* **309**, 314–317 (2005).
- Kornegay, J. N., Tuler, S. M., Miller, D. M. & Levesque, D. C. Muscular dystrophy in a litter of golden retriever dogs. *Muscle Nerve* **11**, 1056–1064 (1988).
- Sharp, N. J. *et al.* An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. *Genomics* **13**, 115–121 (1992).
- Harper, S. Q. *et al.* Modular flexibility of dystrophin: implication for gene therapy of Duchenne muscular dystrophy. *Nature Med.* **8**, 253–261 (2004).
- Minasi, M. G. *et al.* The meso-angioblast: a multipotent, self-renewing cell that originates from the dorsal aorta and differentiates into most mesodermal tissues. *Development* **129**, 2773–2783 (2002).
- Galvez, B. G. *et al.* Complete rescue of dystrophic muscle by mesoangioblasts with enhanced migratory ability. *J. Cell Biol.* **174**, 231–243 (2006).
- Kornegay, J. N., Cundiff, D. D., Bogan, D. J., Bogan, J. R. & Okamura, C. S. The cranial sartorius muscle undergoes true hypertrophy in dogs with golden retriever muscular dystrophy. *Neuromuscul. Disord.* **13**, 493–500 (2003).
- Galli, D. *et al.* Mesoangioblasts, vessel-associated multipotent stem cells, repair the infarcted heart by multiple cellular mechanisms: a comparison with bone marrow progenitors, fibroblasts, and endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* **25**, 692–697 (2005).
- Childers, M. K. *et al.* Skinned single fibers from normal and dystrophin-deficient dogs incur comparable stretch-induced force deficits. *Muscle Nerve* **31**, 768–771 (2005).
- Pavlati, G. K. Regulation of class I MHC expression in skeletal muscle: deleterious effect of aberrant expression on myogenesis. *J. Neuroimmunol.* **125**, 42–50 (2002).
- Krampera, M. *et al.* Bone marrow mesenchymal stem cells inhibit the response of naive and memory antigen-specific T cells to their cognate peptide. *Blood* **101**, 3722–3729 (2003).
- Reinhardt, R. L. *et al.* Visualizing the generation of memory CD4 T cells in the whole body. *Nature* **410**, 101–105 (2001).
- Dell'Agnola, C. *et al.* Hematopoietic stem cell transplantation does not restore dystrophin expression in Duchenne muscular dystrophy dogs. *Blood* **104**, 4311–4318 (2004).
- Karpati, G., Gilbert, R., Petrof, B. J. & Nalbantoglu, J. Gene therapy research for Duchenne and Becker muscular dystrophies. *Curr. Opin. Neurol.* **10**, 430–435 (1997).
- Howell, J. M. *et al.* Use of the dog model for Duchenne muscular dystrophy in gene therapy trials. *Neuromuscul. Disord.* **7**, 325–328 (1997).
- Cerletti, M. *et al.* Dystrophic phenotype of canine X-linked muscular dystrophy is mitigated by adenovirus-mediated utrophin gene transfer. *Gene Ther.* **10**, 750–757 (2003).
- Emery, A. E. Clinical and molecular studies in Duchenne muscular dystrophy. *Prog. Clin. Biol. Res.* **306**, 15–28 (1989).
- Ricordi, C. & Strom, T. M. Clinical islet transplantation: advances and clinical challenges. *Nature Rev. Immunol.* **4**, 259–270 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. G. Roncarolo for helpful discussions; J. Chamberlain for the gift of the lentiviral vector expressing human microdystrophin; C. Drougard for technical assistance; X. Cauchois, I. Gruyer, S. Kouamé, E. Wembe and A. Brindejont and M. Carré at the Centre d'Elevage du Domaine des Souches for their care of the dogs; and N. Borenstein for the systemic delivery of cells. M.S. and G.C. thank P. Luban for a donation. This work was supported by grants from the Association Française contre les Myopathies, Telethon, the Muscular Dystrophy Association, the Parent Project Onlus, the European Community 'Eurostemcell', the Cariplo Foundation and the Italian Ministries of Health and Research. B.G.G. was supported by a 3+3 fellowship from the Centro Nacional de Investigaciones Cardiovasculares, Spain.

Author Contributions M.S. coordinated the work on cells with R.T. and M.G.C.D.; S.B. coordinated the work on dogs with N.G., J.L.T. and I.B.; R.B. and G.D.A. coordinated the physiology experiments with O.P., C.R. and P.M., who also developed, with S.M., the instrument to measure dog tetanic force; A.I. did the histology work; B.G.G. performed the homing experiments; L.P. and M.S. conducted the western blot analysis; M.G. did the immunology experiment; Y.T. and C.B. evaluated the clinical aspects of the work; G.C. coordinated the whole project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.G. (cossu.giulio@hsr.it) or R.B. (roberto.bottinelli@unipv.it).

ARTICLES

An RNA map predicting Nova-dependent splicing regulation

Jernej Ule^{1,2,*†}, Giovanni Stefani^{1,2,*†}, Aldo Mele^{1,2}, Matteo Ruggiu^{1,2}, Xuning Wang³, Bahar Taneri^{4†}, Terry Gaasterland^{4†}, Benjamin J. Blencowe⁵ & Robert B. Darnell^{1,2}

Nova proteins are neuron-specific alternative splicing factors. We have combined bioinformatics, biochemistry and genetics to derive an RNA map describing the rules by which Nova proteins regulate alternative splicing. This map revealed that the position of Nova binding sites (YCA Y clusters) in a pre-messenger RNA determines the outcome of splicing. The map correctly predicted Nova's effect to inhibit or enhance exon inclusion, which led us to examine the relationship between the map and Nova's mechanism of action. Nova binding to an exonic YCA Y cluster changed the protein complexes assembled on pre-mRNA, blocking U1 snRNP (small nuclear ribonucleoprotein) binding and exon inclusion, whereas Nova binding to an intronic YCA Y cluster enhanced spliceosome assembly and exon inclusion. Assays of splicing intermediates of Nova-regulated transcripts in mouse brain revealed that Nova preferentially regulates removal of introns harbouring (or closest to) YCA Y clusters. These results define a genome-wide map relating the position of a *cis*-acting element to its regulation by an RNA binding protein, namely that Nova binding to YCA Y clusters results in a local and asymmetric action to regulate spliceosome assembly and alternative splicing in neurons.

Genome-wide identification of DNA and RNA regulatory elements has provided insights into the combinatorial nature of regulatory networks^{1–7}. However, genome-wide rules relating the position of regulatory elements to their differential activity have not been defined. For example, bioinformatic studies have identified several *cis*-acting RNA motifs that regulate splicing^{3,5–7}, but the *in vivo* relationship between the position of these motifs in pre-mRNA and the activity of RNA-binding proteins that recognize them is limited^{4,8–10}. This issue is of great interest given the role of splicing regulation in tissue-specific functions, particularly in neurons^{11,12}.

The first tissue-specific splicing factors described in vertebrates were the neuronal RNA-binding proteins Nova1 and Nova2 (refs 13–15). Clusters of an RNA motif, YCA Y (where Y indicates a pyrimidine), defined by *in vitro* RNA selection and X-ray crystallography^{16–18}, were necessary and sufficient to confer Nova-dependent regulation on three alternatively spliced target transcripts^{14,19,20}. Recently we developed two technologies, cross-linking and immunoprecipitation (CLIP)²¹ and a splicing microarray²², that identified over 50 Nova-regulated exons; these were validated in *Nova1*^{−/−} or *Nova2*^{−/−} mouse brain. Here we find that the distribution of YCA Y clusters in Nova target pre-mRNAs delineates an RNA map that predicts Nova's action on a genome-wide level, and we relate this RNA map to Nova's effect on splicing.

An RNA map of Nova splicing elements

Sequences of pre-mRNAs containing 48 Nova-regulated exons^{14,20–22} (Supplementary Tables 1 and 2) were enriched in YCA Y motifs near Nova-regulated splice junctions (Fig. 1 and Supplementary Fig. 2). Most YCA Y motifs were closely spaced in clusters averaging 28 nucleotides (90%; Supplementary Fig. 3a, b) and were well conserved

between mouse and human (97%; Supplementary Fig. 3c). Two-thirds (31 out of 48) of Nova-regulated pre-mRNAs contained conserved YCA Y clusters near one of the splice sites, compared with 0.9% of control pre-mRNAs (Supplementary Fig. 4f and Supplementary Tables 1 and 2).

The position of YCA Y clusters correlated with Nova's action in splicing regulation (Supplementary Figs 4 and 5). Among the five major locations of YCA Y clusters, two predict splicing enhancers (correlating with Nova-dependent exon inclusion; termed Nova intronic splicing enhancers NISE2 and NISE3) and three predict splicing silencers (correlating with Nova-dependent exon skipping; termed Nova intronic splicing silencers NISS1 and NISS2 and Nova exonic splicing silencers (subdivided into NESS1 and NESS2); Fig. 1). The two main Nova splicing enhancers were intronic, located downstream of the alternative exon (NISE2: *n* = 18, 118-fold higher abundance than at the corresponding position in control pre-mRNAs, *P* < 0.001 by two-tailed *t*-test, unequal variance; NISE3: *n* = 12, 119-fold higher abundance, *P* = 0.002). These were further classified as having one (12 out of 21, Supplementary Figs 5b, c and 7 and Supplementary Table 1) or both (9 out of 21, Fig. 1c and Supplementary Table 1) NISE2 and NISE3 elements. In addition, three pre-mRNAs had enhancer YCA Y clusters upstream of the alternative exon, in a position analogous to the previously defined NISE1 enhancer element in GlyRα2 pre-mRNA¹⁶ (Fig. 1b; see also Supplementary Fig. 7 and Supplementary Table 2). Nova splicing silencers were either immediately upstream of (NISS2; *n* = 6, 25-fold higher abundance, *P* < 0.05) or within (NESS1, NESS2; *n* = 8, 37-fold higher abundance, *P* = 0.01) the alternative exon, or near the upstream constitutive exon (NISS1; *n* = 5, 70-fold higher abundance, *P* = 0.03; Supplementary Fig. 6a). As a whole, we found five major and two minor positions of YCA Y clusters, three of which

¹Howard Hughes Medical Institute, ²Laboratory of Molecular Neuro-Oncology, ³Biocomputing, Information Technology, and ⁴Laboratory of Computational Genomics, The Rockefeller University, 1230 York Ave, New York, New York 10021, USA. ⁵Banting and Best Department of Medical Research, University of Toronto, Ontario M5G 1L6, Canada. [†]Present addresses: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK (J.U.); Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA (G.S.); UCSD, Scripps Institution of Oceanography, Scripps Genome Center, 8750 Biological Grade, La Jolla, California 92037, USA (B.T., T.G.); Eastern Mediterranean University, Famagusta, Cyprus (B.T.).

*These authors contributed equally to this work.

mirrored positions seen in previously analysed Nova-regulated transcripts^{16,19,20}. We refer to the whole set of positions where YCAY clusters act as Nova-dependent splicing elements as an RNA map (Fig. 1a).

Prediction of Nova-dependent splicing regulation

To test whether the RNA map can predict Nova-dependent splicing regulation *de novo*, we calculated a net YCAY cluster score by subtracting Nova silencer from enhancer cluster scores (Supplementary Fig. 4). Using a stringent scoring method ($|\text{net YCAY cluster score}| > 2.7$), we identified 51 candidate Nova-regulated alternative exons in a genomic database of bioinformatically predicted alternative exons (B.T. and T.G., personal communication). Ten previously validated Nova-regulated exons were among these top predictions. Of the remaining 41 predicted exons, an additional 20 showed significant ($P < 0.01$) splicing change in mouse brain from

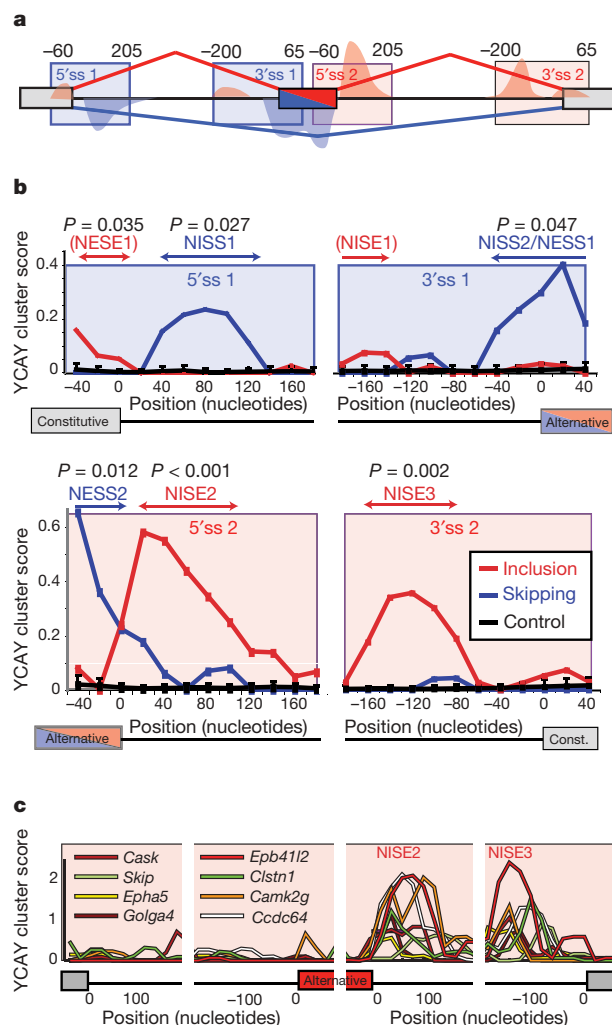


Figure 1 | Definition of the Nova-RNA binding map. **a**, A generic pre-mRNA showing the four regions that define the Nova-RNA binding map (the start and end of each region is labelled by a nucleotide distance to the splice site). Peaks demonstrate the positions of Nova-dependent splicing enhancers (red) or silencers (blue). **b**, A conserved YCAY cluster score (y axis) was calculated as described in Supplementary Methods. The x axis shows the nucleotide position of the centre of the sequence window. At each significant peak of YCAY cluster enrichment in Nova-regulated versus control pre-mRNAs, the P -value (two-tailed t -test, unequal variance) of YCAY clusters is shown. The error value of the control pre-mRNAs represents standard deviation of the mean values of 100 random groups of 20 control pre-mRNAs. **c**, YCAY cluster distribution in eight Nova-regulated pre-mRNAs that contain both NISE2 and NISE3 elements.

Nova1^{-/-} *Nova2*^{-/-} relative to wild-type littermates (Fig. 2; see also Supplementary Fig. S6, Supplementary Table 3 and <http://splicing.rockefeller.edu/map/clusters>).

The RNA map correctly predicted the direction of splicing change for 30 out of 30 Nova-regulated exons (14 exons with a score below -2.7 were downregulated, and 16 exons with a score above 2.7 were upregulated; Fig. 2a), including regulated alternative splice sites and alternative terminal exons (Fig. 2b, c; see also Supplementary Figs 6d, e, 7b). We also assessed some predicted YCAY clusters with absolute scores lower than 2.7 ; although the number of false-positive splicing predictions increased, three additional clusters predicted true Nova-regulated exons, again with correct prediction of splicing change (Supplementary Table 3). Thus, the RNA map accurately predicts splicing silencer and enhancer elements regulated by Nova in mouse brain.

Effect of Nova binding to a splicing silencer

The RNA map suggests a relationship between the position of YCAY clusters and the mechanism of Nova action. We examined the effects of recombinant Nova1 (Supplementary Fig. 12b) on splicing of two different *in vitro* transcribed RNA substrates (Fig. 3a and Supplementary Fig. 12a) harbouring a YCAY cluster (NESS2 element) within the alternative exon 4 (E4) of *Nova1* RNA²⁰, which has a negative net YCAY score (-1.1 , Supplementary Tables 1 and 2). E4 splicing was efficient without Nova, but was inhibited in a dose-dependent manner in the presence of Nova, and this effect was YCAY-dependent (that is, lost when the YCAY motifs were mutated to YAAY; Fig. 3a and Supplementary Fig. 12c, d). These experiments confirm previous data²⁰, and extend them by demonstrating that Nova directly binds to YCAY clusters to regulate splicing according to the rules of an RNA map.

We analysed the effects of Nova on the stepwise formation of the complexes that lead to spliceosome assembly. Nova inhibits formation of complex E, which results from association of U1 small nucleolar ribonucleoprotein (snRNP) and accessory factors that commit pre-mRNA to splicing²³, and this inhibition required intact YCAY clusters (Fig. 3b). Nova also affected the rate of migration of the early pre-spliceosomal complex H in a YCAY-dependent manner (Fig. 3b), which may result from changes in the ribonucleoprotein complex composition or conformation²⁴. To explore whether Nova competes with other proteins assembling on complex H, which itself can be crucial for splicing regulation²⁵, we performed electrophoretic mobility shift assays using nuclear extracts and *Nova1* pre-mRNA substrate (I-E_{YCAY}-I, where I stands for intron, and E for exon). Two shifted bands migrated in a tighter doublet in the presence of Nova (Fig. 3c), and ultraviolet crosslinking studies identified one major protein for which binding was inhibited by *Nova1* (Supplementary Fig. 12e). This protein, identified using biotinylated I-E_{YCAY}-I RNA and mass spectrometry of captured proteins, is hnRNP K (Fig. 3d). Comparing the effects of Nova and hnRNP K on E4 splicing using a transfected minigene harbouring the I-E_{YCAY}-I element²⁰ revealed that Nova had an eightfold greater effect on splicing than hnRNP K (Fig. 3e; see also Supplementary Fig. 12f). These observations suggest that Nova interacts with pre-mRNA before snRNP binding to alter the composition of the pre-spliceosome complex.

Nova can inhibit splicing by blocking U1 snRNP binding

The ability of Nova to interfere with complex E formation suggests that Nova might inhibit access of U1 snRNP to the 5' splice site. We directly tested the efficiency of U1 snRNP recognition of the 5' splice site by crosslinking U1 snRNA to I-E_{YCAY}-I RNA²⁶. A single cross-linked product was detected in the presence of nuclear extract and found to include U1 snRNP (Fig. 3f, lane 3 versus lane 5). Addition of Nova decreased the formation of the U1-RNA complex 2.5-fold in a YCAY-dependent manner (Fig. 3f, lane 4 versus lane 9, Supplementary Fig. 12g). Thus, Nova binding to the NESS2 YCAY cluster hinders U1 snRNP recognition of the 5' splice site.

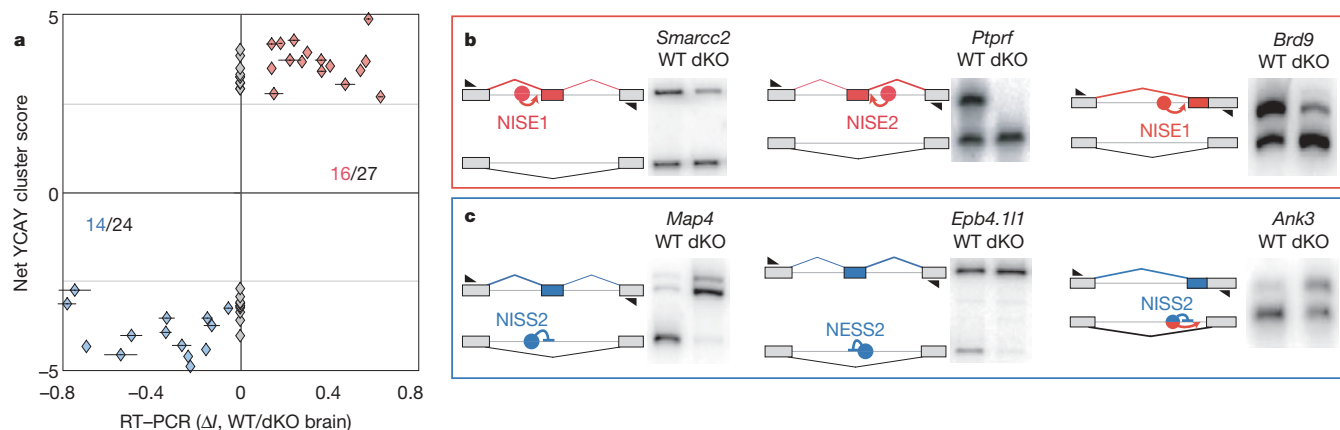


Figure 2 | YCAY cluster position predicts Nova-dependent splicing regulation. **a**, A diagram comparing the bioinformatic prediction of 51 alternative exons having a |net YCAY cluster score| ≥ 2.7 (y axis, $n = 3$, error bars indicate s.d.) with change in the fraction of exon inclusion (Δ)²² as determined by RT-PCR analysis of wild-type (WT) and *Nova1*^{-/-} *Nova2*^{-/-} (double knockout, dKO) mouse brain (x axis). Each RT-PCR experiment was done in biological triplicate (see <http://splicing.rockefeller.edu/map/>). **b**, **c**, Examples of Nova-dependent

inclusion (**b**) or skipping (**c**) of alternative exons or splice sites. In the *Ank3* gene, the same element is in a position to act as an NISS2 on the proximal splice site and an NISE1 on the distal splice site. Colour-coded objects present on the left-hand side of each gel represent the pre-mRNA; black triangles indicate the position of the primers used for RT-PCR; circles represent Nova bound to YCAY clusters; arrows indicate enhanced splice sites; bars indicate silenced splice sites.

We also analysed the effect of Nova on spliceosome assembly after complex E formation. Nova entirely inhibited formation of complexes B and C, with most RNA retained in complex H (Fig. 3g, lane 1 versus lane 2; see also Supplementary Fig. 12h). In the presence of Nova a complex co-migrating with complex A was observed, albeit at a slightly reduced level, which seems to conflict with the observation that Nova inhibits complex E formation (which contains U1 snRNP and precedes the formation of complex A). Although complex A normally contains U1 and U2 snRNPs, atypical complexes have been reported to form on substrates containing only a branch site and polypyrimidine tract and in extracts depleted of U1 snRNP^{27,28}, which agrees with our observation that the band co-migrating with complex A is unaffected by depletion of U1 snRNP but is abolished by depletion of U2 snRNP (Fig. 3g; see also Supplementary Fig. 12i, j). The precise mechanism by which Nova blocks U1 snRNP binding is unknown, but it may involve changes in pre-mRNA secondary structure²⁹ (Supplementary Fig. 12i). Although Nova did not affect complex A formation in a U1-depleted extract (Fig. 3g, lane 4), the Nova-dependent change in complex H migration occurred in the absence of either U1 or U2 snRNP (Fig. 3g, lanes 4 and 6). These results are consistent with Nova acting locally to change the composition of the pre-spliceosome complex, thereby interfering with the binding of U1 (but not U2) snRNP, and inhibiting exon inclusion.

Effect of Nova binding to a splicing enhancer

To explore the mechanism by which Nova acts on an intronic YCAY cluster, we assayed *in vitro* splicing of a substrate (E-I-E-YCAY-E RNA), modelled after GABA_A (γ -aminobutyric acid) receptor γ 2L pre-mRNA¹⁹ (positive net YCAY score = 3.2; Supplementary Tables 1 and 2), that harboured a YCAY cluster in the middle of a short intron at a position analogous to NISE2 and NISE3 elements. Nova reproducibly mediated a >2-fold increase in the ratio of exon included to skipped product and an accumulation of the second intron lariat; this effect was YCAY-dependent (Fig. 4a; see also Supplementary Fig. 13a–c). When we analysed Nova's effect on spliceosome assembly using an RNA substrate in which the first two exons were joined (EE-I-YCAY-E; Fig. 4b) to ensure that the transcript assembled into a single spliceosome, Nova again mediated an increase in the spliced product (1.7-fold, $n = 4$, $P < 1.2 \times 10^{-3}$, two-tailed t -test, unequal variance; Supplementary Fig. 13). This effect correlated with increased formation of complex A, followed by complexes B and C (Fig. 4c, d; 1.6-fold, $n = 4$, $P < 0.04$);

additional studies suggest that Nova might act on this substrate to affect RNA folding (Supplementary Fig. 13e). These results demonstrate that Nova action at a NISE YCAY cluster promotes spliceosome assembly.

Prediction of asymmetric Nova action on intron removal

Our *in vitro* analysis of Nova-dependent inhibition of exon inclusion suggested that Nova acts locally to affect spliceosome assembly, as Nova inhibited U1 snRNP binding to the E_{YCAY}-I-E RNA 5' splice site (where the YCAY cluster is located) but not U2 snRNP binding to the 3' splice site of the same intron (Fig. 3f, g; see also Supplementary Fig. 12i, j). To relate this observation to Nova-dependent splicing regulation in mouse brain, we assessed Nova's effect on splicing of the two introns flanking the alternative exons. We measured RNA splicing intermediates (with one intron unspliced and the other exon–exon junction fully spliced) in 30 exons containing or flanked by Nova silencer or enhancer elements, comparing RNA from *Nova1*^{-/-} *Nova2*^{-/-} relative to wild-type mouse brain (Fig. 5; see also Supplementary Fig. 14).

We tested 16 transcripts in which Nova enhanced exon inclusion via YCAY clusters in the downstream intron (NISE2 or NISE3) and found that Nova was associated with a 5- to 50-fold increase in the 3' splicing intermediate lacking the intron 3' of the alternative exon (Fig. 5b; see also Supplementary Fig. 14b). In contrast, Nova was rarely associated with any change in the 5' splicing intermediate lacking the intron 5' of the alternative exon, and when it was, the exons were short (~25 nucleotides) and the effects smaller than on the intron with YCAY clusters (Supplementary Table 3). We also tested two transcripts containing YCAY clusters in the intron upstream of the alternative exon (NISE1), and found that in one case (*Ddr1*, Fig. 5a; see also Supplementary Fig. 14a), Nova splicing enhancement was associated with an asymmetric increase (fourfold) in the 5', but not 3', splicing intermediate. Taken together, we detected a significant Nova-dependent increase ($P < 0.05$, two-tailed t -test, unequal variance) in at least one splicing intermediate in 17 out of 18 transcripts where Nova promoted alternative exon inclusion. In 12 of these, a significant increase was evident only in a single splicing intermediate, corresponding in each case to an asymmetric action on splicing of the intron harbouring the YCAY cluster (Fig. 5a, b; see also Supplementary Fig. 14a, b and Supplementary Table 3).

We also analysed Nova's effect on splicing intermediates in transcripts where Nova inhibited alternative exon inclusion. Because

Nova's silencing action on splicing intermediates seemed to differ when clusters were located at the 5' or 3' end of the alternative exon, we considered separately exonic splicing silencers within 60 nucleotides of either the 3' splice site (NISS2, NESS1) or the 5' splice site (NESS2). Nova splicing inhibition was associated with a significant decrease in splicing of the intron harbouring YCAY clusters in 7 out of 7 Nova-regulated exons with upstream clusters and in 4 out of 4 with downstream clusters (NISS2 and NESS1, 2- to 16-fold; NESS2, 2- to 18-fold, respectively, $P < 0.05$, two-tailed t -test, unequal vari-

ance; Fig. 5c, d, Supplementary Fig. 14d–f and Supplementary Table 3). With the exception of two transcripts containing additional YCAY clusters (Supplementary Fig. 14g), the effect of Nova was limited in each case to one splicing intermediate. Taken together, of the 30 cases tested, 19 had significant ($P < 0.05$) Nova-dependent effects in only one splicing intermediate, and these all showed an asymmetric effect of Nova on splicing of the intron harbouring (or closest to) the YCAY cluster (Supplementary Table 3).

Discussion

We used over 50 previously identified, genetically validated Nova-regulated alternative transcripts^{14,20–22} to derive a general RNA map relating the sequence and position of YCAY clusters to Nova activity (Fig. 1). When used to screen the mouse genome, this RNA map correctly predicted Nova splicing silencers and enhancers in another 30 alternatively spliced transcripts (Fig. 2). The predictive accuracy of the map suggested a mechanistic link between the position where Nova binds YCAY clusters and the splicing outcome. Assays in a reconstituted splicing system demonstrated that Nova acts on a NESS2 YCAY cluster to interfere specifically with U1 snRNP binding

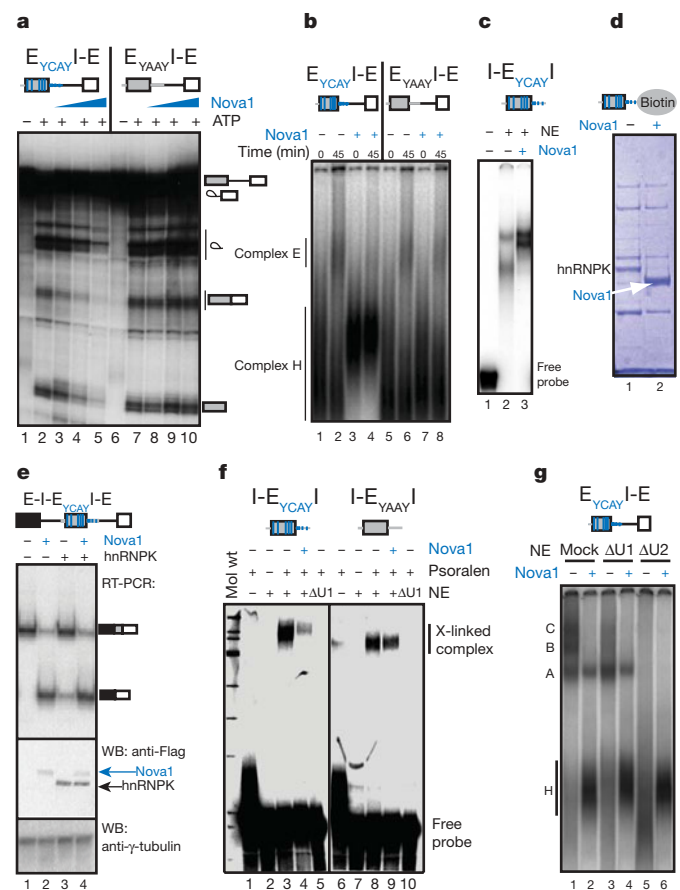


Figure 3 | Nova inhibits splicing of an RNA substrate containing a NESS2 element. **a**, In vitro splicing assay using E_{YCAY}-I-E (lanes 1–5) or E_{YAAY}-I-E (lanes 6–10) RNA (1 nM) and nuclear extract in the absence or presence of ATP. Recombinant Nova1 was included in concentrations ranging from 0.03 to 0.12 μM. **b**, Nova inhibits the early phase of spliceosome assembly. A 1 nM concentration of E_{YCAY}-I-E (lanes 1–4) or E_{YAAY}-I-E (lanes 5–8) RNA was incubated on ice or at 30 °C in the absence of ATP. Nova was included at 0.03 μM concentration. **c**, Electrophoretic mobility shift assay. Labelled I-E_{YCAY}-I (1 nM) was incubated in buffer (lane 1) or in HeLa nuclear extract (lane 2). Nova (0.12 μM) was included in lane 3. **d**, RNA affinity purification was performed using 3'-biotinylated I-E_{YCAY}-I from HeLa nuclear extract without (lane 1) or with recombinant Nova (lane 2). Mass spectrometry analysis identified the 58-kDa band as hnRNP K. **e**, Effects of Flag-HA-Nova1 and Flag-hnRNP K overexpression on splicing of empty expression vector (lane 1) or E-I-E_{YCAY}-I-E minigene (lanes 2–4) (Supplementary Fig. 11a). **f**, Psoralen crosslinking assay. Labelled I-E_{YCAY}-I (lanes 1–5) or I-E_{YAAY}-I RNA (lanes 6–10) was incubated at 1 nM concentration in the absence (lane 1 and 6) or presence (lanes 2–5 and 7–10) of HeLa nuclear extract (NE). In lanes 5 and 10, the 5' end of U1 snRNA was removed by oligodeoxynucleotide-mediated RNase H treatment (ΔU1). Recombinant Nova1 (0.12 μM final concentration) was included in lanes 4 and 9. **g**, Spliceosome assembly in the presence of ATP. Labelled E_{YCAY}-I-E pre-mRNAs (1 nM) were incubated in nuclear extracts (NE) that were mock-depleted (lanes 1 and 2), depleted of U1 snRNP (ΔU1, lanes 3 and 4), or U2 snRNP (lanes 5 and 6). Recombinant Nova (0.03 μM) was included in lanes 2, 4 and 6. Vertical blue lines in all cartoons represent YCAY motifs.

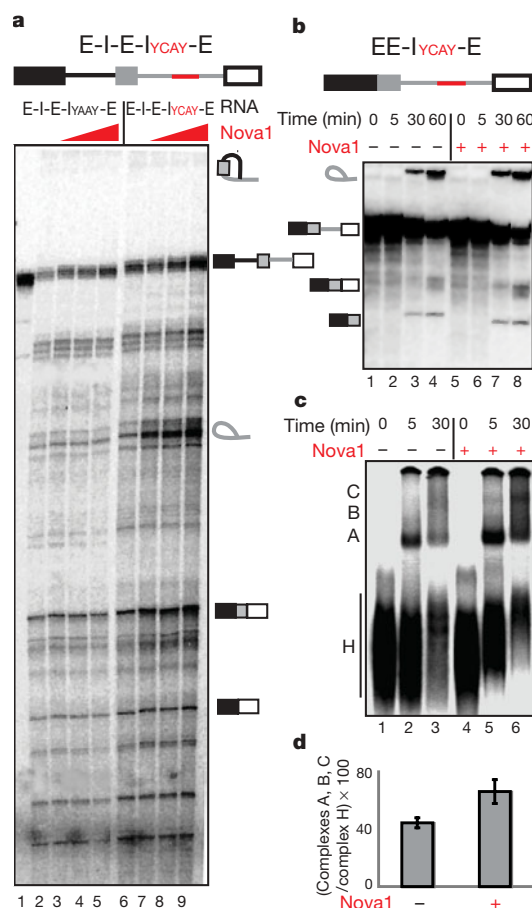


Figure 4 | Nova binding to a splicing enhancer promotes spliceosome assembly. **a**, Labelled E-I-E-I_{YAAY}-E (lanes 1–5) or E-I-E-I_{YCAY}-E (lanes 6–9) RNA was incubated for 4 h at 30 °C with HeLa nuclear extract in the absence of ATP (lane 1) or in splicing conditions (lanes 2–9) at 1 nM concentration in the absence (lanes 2 and 6) or presence of Nova (concentrations ranging from 0.02 to 0.08 μM, lanes 3–5 and 7–9). **b**, Labelled EE-I_{YCAY}-E RNA was incubated in splicing conditions in the absence (lanes 1–4) or presence (lanes 5–8) of 0.12 μM Nova. **c**, Spliceosome assembly in the presence of ATP. Radiolabelled EE-I_{YCAY}-E (1 nM) was incubated in nuclear extract (NE) in the absence or presence of recombinant Nova (0.12 μM). **d**, Quantification of the gel shown in **c** ($n = 4$, error bars indicate s.d.).

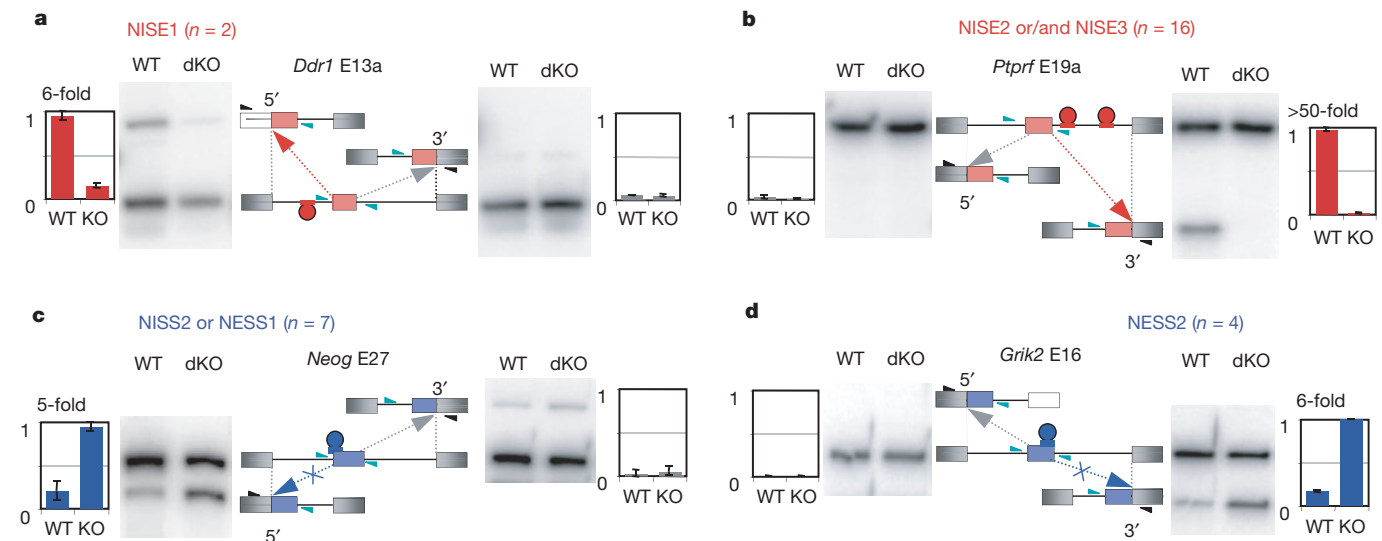


Figure 5 | Analysis of splicing intermediates in Nova-regulated RNAs. Quantification of splicing intermediates that contain one unspliced intron and the other exon–exon junction spliced, using RNA from *Nova1*^{−/−}*Nova2*^{−/−} (dKO) and wild-type littermate brains. In addition to the primer pair that amplifies the splicing intermediate, two additional primers amplifying pre-mRNA were used to monitor the efficiency of all primers (for visual clarity, only one pre-mRNA control band is shown here; both control bands are shown in Supplementary Fig. 14). The graphs show quantification of bands representing splicing intermediates, normalized to the pre-mRNA bands ($n = 3$, error bars indicate s.d.). Because the two splicing intermediates represent two alternative pathways leading to exon inclusion, the added value of bands in wild-type samples was normalized to 1 when

Nova promotes exon inclusion, and the added value in *Nova1*^{−/−}*Nova2*^{−/−} double knockout samples was normalized to 1 when Nova promotes exon skipping. Colour-coded objects present represent the pre-mRNA, splicing intermediates, the position of the primers used for RT–PCR (triangles), and position of Nova bound to YCAAY clusters (circle). Each RT–PCR experiment was done in biological triplicate and the identity of bands was verified by sequencing, and where PCR products were short enough, quantified by real-time PCR (Supplementary Fig. 14). **a**, Representative of two tested pre-mRNAs containing a NISE1 element. **b**, Representative of 16 tested pre-mRNAs containing a NISE2 and/or NISE3 elements. **c**, Representative of seven tested pre-mRNAs containing a NESS1 or NISS2 elements. **d**, Representative of four pre-mRNAs containing a NESS2 element.

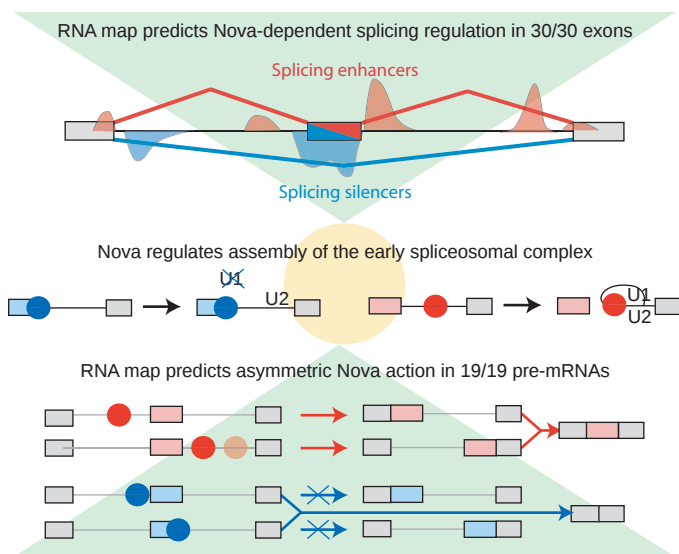


Figure 6 | An RNA map providing a comprehensive view of Nova function. An RNA map was defined using an algorithm that analysed the distribution of Nova binding sites (YCAAY clusters)^{14–16} in genetically validated Nova-target pre-mRNAs^{12,18–20}. The RNA map was applied to a genome-wide screen to predict correctly whether Nova acted to enhance or block alternative exon inclusion in all of the top scoring and validated 30 exons (top triangle). The RNA map relates to the precise mechanism of Nova action to enhance or inhibit spliceosome assembly in the vicinity of YCAAY clusters, as determined in a reconstituted system *in vitro* (circle). These results were generalized by quantification of splicing intermediates in Nova knockout brain, which showed that the RNA map predicts the *in vivo* site of Nova action in each of 19 pre-mRNAs where we detected an asymmetry in Nova's action on removal of the two introns flanking the alternative exons (bottom triangle). This data defines a genome-wide RNA map in which the position of *cis*-acting elements determines the outcome of regulation by the splicing factor Nova.

(Fig. 3) and on a NISE2/3 YCAAY cluster to promote spliceosome assembly (Fig. 4). Furthermore, quantification of splicing intermediates in brain of Nova knockout (*Nova1*^{−/−}*Nova2*^{−/−}) mice demonstrated a direct and asymmetric *in vivo* action of Nova on introns harbouring (or closest to) YCAAY binding elements (Fig. 5). These data support a mechanistic model in which the RNA map dictates a local action of Nova to regulate spliceosome assembly (Fig. 6 and Supplementary Fig. 1).

A validated RNA map provides new insights into functional RNA–protein interactions. Defining Nova binding sites as YCAAY clusters^{16–21} (Supplementary Fig. 3) allowed us to develop a bioinformatic procedure with high predictive accuracy, suggesting more generally that searching for clustered motifs may uncover regulatory elements overlooked in studies focusing on single, short splicing motifs^{3,6,7,10}. The degree of YCAAY cluster conservation between mouse and human genomes (97%; Supplementary Fig. 3c) agrees with previous bioinformatic estimates that 98% of major form alternative exons are conserved³⁰. Nova-regulated targets that did not have high YCAAY cluster scores within the region of the current RNA map (Supplementary Fig. 4 and Supplementary Table 1) may be regulated by YCAAY clusters located outside of the current RNA map, by Nova-dependent regulation of other splicing factors, by interactions of Nova with other splicing factors, or by other means. Notably, we found an over-represented motif, CACCA, in nine NESS elements with top YCAAY cluster scores. This pentamer partially overlaps Nova binding sites (YCACCA; Supplementary Fig. 9b) and was previously shown to act as an exonic enhancer^{31,32}, suggesting that in some transcripts Nova may compete or synergize with other factors (such as hnRNP K (Fig. 3) or brPTB/nPTB³³) to regulate splicing.

Nova's displacement of U1 snRNP by binding to the NESS2 element (Figs 3 and 5c) is reminiscent of the ability of Sxl³⁴, hnRNP A1^{35,36} and PTB^{25,37} to displace U2AF₆₅ or U2 snRNP, which is how Nova might act when binding NISS2 and NESS1 elements in the vicinity of the 3' splice site (Fig. 5d). In contrast, Nova binding to

intronic YCAY clusters generally promotes spliceosome assembly. Although splicing enhancers have most often been associated with exonic sequences^{3,5,8,31}, in some instances proteins such as TIA1³⁷, KSRP, hnRNP F/H^{10,38}, Fox1 and Fox2³⁹ promote spliceosome assembly by binding to intronic sequences downstream of the exon. Notably, Nova-regulated exons flanked by NISE2 elements are generally shorter than 50 nucleotides ($n = 12$, $P = 0.003$; Supplementary Table 3 and Supplementary Fig. 9a), consistent with suggestions that intronic enhancer elements are particularly important in regulating short exons⁴⁰, which themselves have a lower chance of harbouring exonic splicing enhancers. Of the 13 Nova-target pre-mRNAs with NISE3 elements (near the downstream constitutive exon), 10 also contain a NISE2 element (near the alternative exon; Fig. 1c and Supplementary Tables 1 and 3), suggesting that Nova might bind both sites and multimerize⁴¹ to form an RNA loop bringing the 5' splice site and the branch site into close proximity. Models of splicing regulation by factors involved in RNA looping or binding to multiple sites have been proposed for brPTB/nPTB⁴², hnRNP A/B^{36,43} and hnRNP F/H⁴³.

We quantified partially spliced RNAs in the brain of *Nova1*^{-/-} *Nova2*^{-/-} mice, demonstrating that Nova in most cases (19 out of 30) acts asymmetrically to regulate splicing of only one intron flanking the alternative exon. In all of these cases (19 out of 19) the regulated intron contains (or is proximal to) YCAY clusters (Fig. 6 and Supplementary Fig. 1). These observations support a model in which local, asymmetric actions of Nova enhance splicing of the 3' and 5' intron via NISE1 and NISE2/NISE3, respectively, and silence splicing of the 3' and 5' intron via NISS2/NESS1 and NESS2, respectively.

How can exon skipping be induced by Nova if it only prevents splicing of one intron flanking the alternative exon, while splicing of the other intron remains equally efficient? Intuitively, exon skipping should require repression of both splicing events to generate a product in which the exon is excluded. However, pre-mRNAs containing exons that are silenced in the presence of Nova most often display some degree of asymmetry of splicing intermediates even in the absence of Nova, suggesting that splicing of one of the introns represents the rate limiting step for exon inclusion (Fig. 5c, d; see also Supplementary Fig. 14c–e). The mechanism underlying this pre-determined asymmetry is unclear, because it does not correlate with splice site scores of Nova-regulated exons, which on average have consensus splice site sequences similar to those of constitutive exons (Supplementary Fig. 10). Notably, in all transcripts with such pre-determined asymmetry, the Nova silencer elements are located at the intron that is normally removed first, thus enabling Nova to inhibit the rate-limiting step for exon inclusion (Fig. 5c, d; see also Supplementary Fig. 14c–e). Our data support previous biochemical studies indicating that pre-mRNAs often follow a preferential splicing order^{44–47}, and suggest that such an order may allow proteins like Nova to change splicing outcome by locally regulating the rate limiting step leading to alternative exon inclusion.

This study identified 23 new Nova-regulated exons (Supplementary Table 3), 12 of which encode proteins with known functions in brain. Eleven out of 12 of these proteins act in adhesion, cytoskeleton binding, neuronal inhibition, signalling, or synapse formation and plasticity (<http://splicing.rockefeller.edu/map/validated>). These observations agree with and extend previous findings that Nova, as well as other regulators of gene expression, tend to co-regulate genes encoding functionally related proteins^{21,22,48} (see ref. 11 for a review). For example, previous microarray analysis had identified *Ptprf* exon 19a, *Ank3* exon 17 and *Epb4.1* exon 14 (ref. 22), and the current study identified *Ptprf* exon 6a, *Ank3* exon 31 and *Epb4.1* exon 16 as Nova targets; in addition, two Nova-regulated exons encode variants of Dab1 (Supplementary Figs 6c, 7a), a protein known to participate in the pathway of reelin-dependent synaptic plasticity modulation⁴⁹. It remains to be tested whether alternative splicing of these and other Nova-regulated exons contributes to long term potentiation of slow

inhibitory postsynaptic currents, which is defective in *Nova2*^{-/-} hippocampus⁵⁰, or other aspects of neuronal physiology.

Our analysis of functional Nova–RNA interactions leads to several new conclusions (Fig. 6 and Supplementary Fig. 1). We provide an RNA map predicting the activity of a splicing regulator in which splicing silencers differ from enhancers in their position on the pre-mRNA, but share a similar sequence. Whereas sequences close to alternative exons have been the focus of previous bioinformatic studies^{3–10}, we find that sequences close to the flanking constitutive exons have equally important roles in the regulation of alternative splicing. Finally, our analysis provides a new model whereby Nova regulates exon inclusion through a direct and asymmetric action on one of the flanking introns, by locally blocking or enhancing assembly of the spliceosome. As a general approach, our data suggest that a detailed understanding of the nature of protein–RNA interactions, together with identification of a large set of RNA targets validated in a genetic system, can be combined with bioinformatics and biochemistry to identify the sites and mechanisms of action of alternative splicing factors.

METHODS

Pre-mRNAs containing 48 exons that were previously found to be Nova-regulated^{14,20–22} (excluding special cases, such as intron retention or redundant consecutive exons, Supplementary Tables 1 and 2) were analysed to optimize the procedure for calculating the YCAY cluster score (Supplementary Fig. 3; see also Supplementary Methods). The YCAY cluster score was quantified using three criteria: the number of YCAY motifs, the distance between YCAY motifs, and the degree of evolutionary conservation, using an algorithm that maximized the accuracy and efficiency of predicting Nova-regulated, but not control exons (Supplementary Fig. 4). Net YCAY cluster score was calculated by subtracting the maximum score of NISS and NESS elements from the maximum score of NISE elements, such that a positive net YCAY cluster score predicts that Nova promotes exon inclusion, and a negative score predicts that Nova promotes exon skipping. Fifty-one candidate alternative exons had a predicted |net YCAY cluster score| > 2.7, and we tested splicing of these in *Nova1*^{-/-} *Nova2*^{-/-} embryonic day 18 mouse brain by polymerase chain reaction with reverse transcription (RT–PCR). All 51 predicted alternative exon candidates were successfully detected and confirmed by sequencing (see <http://splicing.rockefeller.edu/map/>); it is uncertain whether our ability to predict alternative exons and genes that are expressed at day 18 embryonic brain relates to our stringent filtering for alternative exons, the presence of YCAY clusters, or other variables.

Recombinant Nova1 (Supplementary Fig. 12b) was used for all *in vitro* assays, described in Supplementary Methods. The substrates used for *in vitro* assays were E-I-E_{YCAY}-I-E and its short versions (E_{YCAY}-I-E and I-E_{YCAY}-I), and E-I-E-I_{YCAY}-E and its short version E-I_{YCAY}-E. Analysis of splicing intermediates by RT–PCR was performed using RNA from *Nova1*^{-/-} *Nova2*^{-/-} embryonic day 18 mouse brain and quantitative autoradiography or real-time PCR.

Received 5 May; accepted 3 October 2006.

Published online 25 October 2006.

- Hallikas, O. *et al.* Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124**, 47–59 (2006).
- Wasserman, W. W. & Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nature Rev. Genet.* **5**, 276–287 (2004).
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
- Hui, J. *et al.* Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.* **24**, 1988–1998 (2005).
- Liu, H. X., Zhang, M. & Krainer, A. R. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**, 1998–2012 (1998).
- Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
- Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C. B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA* **102**, 2850–2855 (2005).
- Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).
- Ryder, S. P., Frater, L. A., Abramovitz, D. L., Goodwin, E. B. & Williamson, J. R. RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. *Nature Struct. Mol. Biol.* **11**, 20–28 (2004).
- Han, K., Yeo, G., An, P., Burge, C. B. & Grabowski, P. J. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* **3**, e158 (2005).

11. Ule, J. & Darnell, R. B. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.* **16**, 102–110 (2006).
12. Licatalosi, D. D. & Darnell, R. B. Splicing regulation in neurologic disease. *Neuron* **52** (1), 93–101 (2006).
13. Buckanovich, R. J., Yang, Y. Y. & Darnell, R. B. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16**, 1114–1122 (1996).
14. Jensen, K. B. *et al.* Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
15. Yang, Y. Y., Yin, G. L. & Darnell, R. B. The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl Acad. Sci. USA* **95**, 13254–13259 (1998).
16. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell. Biol.* **17**, 3194–3201 (1997).
17. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K. & Darnell, R. B. The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl Acad. Sci. USA* **97**, 5740–5745 (2000).
18. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).
19. Dredge, B. K. & Darnell, R. B. Nova regulates GABA_A receptor $\gamma 2$ alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol. Cell. Biol.* **23**, 4687–4700 (2003).
20. Dredge, B. K., Stefani, G., Engelhard, C. C. & Darnell, R. B. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24**, 1608–1620 (2005).
21. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
22. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nature Genet.* **37**, 844–852 (2005).
23. Michaud, S. & Reed, R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534–2546 (1991).
24. Bennett, M., Michaud, S., Kingston, J. & Reed, R. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev.* **6**, 1986–2000 (1992).
25. Sharma, S., Falick, A. M. & Black, D. L. Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell* **19**, 485–496 (2005).
26. Tarn, W. Y. & Steitz, J. A. SR proteins can compensate for the loss of U1 snRNP functions *in vitro*. *Genes Dev.* **8**, 2704–2717 (1994).
27. Barabino, S. M., Blencowe, B. J., Ryder, U., Sproat, B. S. & Lamond, A. I. Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* **63**, 293–302 (1990).
28. Query, C. C., McCaw, P. S. & Sharp, P. A. A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract. *Mol. Cell. Biol.* **17**, 2944–2953 (1997).
29. Buratti, E. & Baralle, F. E. Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.* **24**, 10505–10514 (2004).
30. Xing, Y. & Lee, C. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. *Nature Rev. Genet.* **7**, 499–509 (2006).
31. Coulter, L. R., Landree, M. A. & Cooper, T. A. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* **17**, 2143–2150 (1997).
32. Gersappe, A. & Pintel, D. J. CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minute virus of mice NS2-specific exon in both singly and doubly spliced mRNAs. *Mol. Cell. Biol.* **19**, 364–375 (1999).
33. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97**, 6350–6355 (2000).
34. Valcarcel, J., Singh, R., Zamore, P. D. & Green, M. R. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171–175 (1993).
35. Del Gatto-Konczak, F., Olive, M., Gesnel, M. C. & Breathnach, R. hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer. *Mol. Cell. Biol.* **19**, 251–260 (1999).
36. Zhu, J., Mayeda, A. & Krainer, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell* **8**, 1351–1361 (2001).
37. Izquierdo, J. M. *et al.* Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell* **19**, 475–484 (2005).
38. Chou, M. Y., Rooke, N., Turck, C. W. & Black, D. L. hnRNP H is a component of a splicing enhancer complex that activates a c-Src alternative exon in neuronal cells. *Mol. Cell. Biol.* **19**, 69–77 (1999).
39. Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25**, 10005–10016 (2005).
40. Carlo, T., Sterner, D. A. & Berget, S. M. An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA* **2**, 342–353 (1996).
41. Lewis, H. A. *et al.* Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Struct. Fold. Des.* **7**, 191–203 (1999).
42. Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H. & Black, D. L. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-Src neural-specific splicing. *Mol. Cell* **5**, 949–957 (2000).
43. Martinez-Contreras, R. *et al.* Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.* **4**, e21 (2006).
44. Gee, S. L. *et al.* Alternative splicing of protein 4.1R exon 16: ordered excision of flanking introns ensures proper splice site choice. *Blood* **95**, 692–699 (2000).
45. Kessler, O., Jiang, Y. & Chasin, L. A. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol.* **13**, 6211–6222 (1993).
46. Lang, K. M. & Spritz, R. A. *In vitro* splicing pathways of pre-mRNAs containing multiple intervening sequences? *Mol. Cell. Biol.* **7**, 3428–3437 (1987).
47. Tsai, M. J., Ting, A. C., Nordstrom, J. L., Zimmer, W. & O'Malley, B. W. Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. *Cell* **22**, 219–230 (1980).
48. Cook, H. L. *et al.* Small nuclear RNAs encoded by *Herpesvirus saimiri* upregulate the expression of genes linked to T cell activation in virally transformed T cells. *Curr. Biol.* **15**, 974–979 (2005).
49. Beffert, U. *et al.* Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. *Neuron* **47**, 567–579 (2005).
50. Huang, C. S. *et al.* Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. *Cell* **123**, 105–118 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Eom and D. Licatalosi for help with breeding *Nova1^{-/-} Nova2^{-/-}* mice and providing mouse brain; P. Ariel for technical assistance; M. Babu, C. Smith, J. Valcarcel, G. Yeo, D. Licatalosi, J. Darnell, S. Xie, and S. W. Chi for critically reading the manuscript; D. Karolchik and J. Jackson for help with the UCSC Genome Bioinformatics tools; J. Okano for the hnRNP K expression construct; A. Krainer and L. Manche for help with *in vitro* splicing assays; K. Dredge for DNA constructs; D. Black for sharing unpublished results; and M. Konarska and members of the laboratory for discussions. Supported by the NIH (R.B.D.) and the Howard Hughes Medical Institute, the tumour immunology program of Cancer Research Institute (J.U.) and a Human Frontiers Science Program Fellowship (M.R.). R.B.D. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions J.U. bioinformatically defined the RNA map, and predicted and analysed Nova-target exons and splicing intermediates; G.S. performed *in vitro* studies of the mechanisms of Nova action and its effects on spliceosome assembly; A.M. purified PCR products for sequencing; M.R. characterized the *Nova1^{-/-} Nova2^{-/-}* mice; X.W. wrote the sequence analysis programs; B.T. and T.G. provided the database of alternative exons; B.J.B. provided depleted extracts; and R.B.D. supervised all studies. The manuscript was prepared by J.U., G.S. and R.B.D., with the participation of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B.D. (darnell@rockefeller.edu).

Decoding the ancient Greek astronomical calculator known as the Antikythera Mechanism

T. Freeth^{1,2}, Y. Bitsakis^{3,5}, X. Moussas³, J. H. Seiradakis⁴, A. Tselikas⁵, H. Mangou⁶, M. Zafeiropoulou⁶, R. Hadland⁷, D. Bate⁷, A. Ramsey⁷, M. Allen⁷, A. Crawley⁷, P. Hockley⁷, T. Malzbender⁸, D. Gelb⁸, W. Ambrisco⁹ & M. G. Edmunds¹

The Antikythera Mechanism is a unique Greek geared device, constructed around the end of the second century BC. It is known^{1–9} that it calculated and displayed celestial information, particularly cycles such as the phases of the moon and a luni-solar calendar. Calendars were important to ancient societies¹⁰ for timing agricultural activity and fixing religious festivals. Eclipses and planetary motions were often interpreted as omens, while the calm regularity of the astronomical cycles must have been philosophically attractive in an uncertain and violent world. Named after its place of discovery in 1901 in a Roman shipwreck, the Antikythera Mechanism is technically more complex than any known device for at least a millennium afterwards. Its specific functions have remained controversial^{11–14} because its gears and the inscriptions upon its faces are only fragmentary. Here we report surface imaging and high-resolution X-ray tomography of the surviving fragments, enabling us to reconstruct the gear function and double the number of deciphered inscriptions. The mechanism predicted lunar and solar eclipses on the basis of Babylonian arithmetic-progression cycles. The inscriptions support suggestions of mechanical display of planetary positions^{9,14,15}, now lost. In the second century BC, Hipparchos developed a theory to explain the irregularities of the Moon's motion across the sky caused by its elliptic orbit. We find a mechanical realization of this theory in the gearing of the mechanism, revealing an unexpected degree of technical sophistication for the period.

The bronze mechanism (Fig. 1), probably hand-driven, was originally housed in a wooden-framed case¹ of (uncertain) overall size 315 × 190 × 100 mm (Fig. 2). It had front and back doors, with astronomical inscriptions covering much of the exterior of the mechanism. Our new transcriptions and translations of the Greek texts are given in Supplementary Note 2 ('glyphs and inscriptions'). The detailed form of the lettering can be dated to the second half of the second century BC, implying that the mechanism was constructed during the period 150–100 BC, slightly earlier than previously suggested¹. This is consistent with a date of around 80–60 BC for the wreck^{1,16} from which the mechanism was recovered by some of the first underwater archaeology. We are able to complete the reconstruction¹ of the back door inscription with text from fragment E, and characters from fragments A and F (see Fig. 1 legend for fragment nomenclature). The front door is mainly from fragment G. The text is astronomical, with many numbers that could be related to planetary motions; the word "sterigmos" (ΣΤΗΡΙΓΜΟΣ, translated as 'station' or 'stationary point') is found, meaning where a planet's apparent motion changes direction, and the numbers may relate to

planetary cycles. We note that a major aim of this investigation is to set up a data archive to allow non-invasive future research, and access to this will start in 2007. Details will be available on www.antikythera-mechanism.gr.

The back door inscription mixes mechanical terms about construction ("trunnions", "gnomon", "perforations") with astronomical periods. Of the periods, 223 is the Saros eclipse cycle (see Box 1 for a brief explanation of astronomical cycles and periods). We discover the inscription "spiral divided into 235 sections", which is



Figure 1 | The surviving fragments of the Antikythera Mechanism. The 82 fragments that survive in the National Archaeological Museum in Athens are shown to scale. A key and dimensions are provided in Supplementary Note 1 ('fragments'). The major fragments A, B, C, D are across the top, starting at top left, with E, F, G immediately below them. 27 hand-cut bronze gears are in fragment A and one gear in each of fragments B, C and D. Segments of display scales are in fragments B, C, E and F. A schematic reconstruction is given in Fig. 2. It is not certain that every one of the remaining fragments (numbered 1–75) belong to the mechanism. The distinctive fragment A, which contains most of the gears, is approximately 180 × 150 mm in size. We have used three principal techniques to investigate the structure and inscriptions of the Antikythera Mechanism. (1) Three-dimensional X-ray microfocus computed tomography²⁴ (CT), developed by X-Tek Systems Ltd. The use of CT has been crucial in making the text legible just beneath the current surfaces. (2) Digital optical imaging to reveal faint surface detail using polynomial texture mapping (PTM)^{25,26}, developed by Hewlett-Packard Inc. (3) Digitized high-quality conventional film photography.

¹Cardiff University, School of Physics and Astronomy, Queens Buildings, The Parade, Cardiff CF24 3AA, UK. ²Images First Ltd, 10 Hereford Road, South Ealing, London W5 4SE, UK. ³National and Kapodistrian University of Athens, Department of Astrophysics, Astronomy and Mechanics, Panepistimiopolis, GR-15783, Zographos, Greece. ⁴Aristotle University of Thessaloniki, Department of Physics, Section of Astrophysics, Astronomy and Mechanics, GR-54124 Thessaloniki, Greece. ⁵Centre for History and Palaeography, National Bank of Greece Cultural Foundation, P. Skouze 3, GR-10560 Athens, Greece. ⁶National Archaeological Museum of Athens, 1 Tositia Str., GR-10682 Athens, Greece. ⁷X-Tek Systems Ltd, Tring Business Centre, Icknield Way, Tring, Hertfordshire HP23 4JX, UK. ⁸Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, California 94304, USA. ⁹Foxhollow Technologies Inc., 740 Bay Road, Redwood City, California 94063, USA.

Box 1 | Astronomical cycles known to the Babylonians

The lunar (or synodic) month is the interval between the Moon being at the same phase—for example, full moon to full moon. The Metonic cycle results from the close equality of 19 years to 235 lunar months. It represents the return to the same phase of the Moon on the same date in the year. After the cycle, the Sun, Moon and Earth are back in nearly the same relative orientations. The Moon appears to return to the same point in the sky relative to the zodiac in a sidereal month, and in 19 years there are $235 + 19 = 254$ sidereal months. The 76-year Callippic cycle is four Metonic cycles minus one day—and improves the accuracy of reconciling solar years with whole numbers of lunar months.

The Saros is an eclipse repeat cycle. If either a solar or lunar eclipse occurs, a very similar eclipse will occur 223 lunar months later²³. A record of past eclipses can thus be used to predict future occurrences. The cycle arises from the coincidence of three orbital periods of the Moon. These are: (1) same phase to same phase, 223 synodic months—eclipses will of course only occur at new or full Moon in the month; (2) the lunar crossing of the Earth–Sun orbital plane, 242 draconitic months—eclipses can only occur near these points (nodes) of co-alignment; (3) similar Earth–Moon distances which occur on the period from apogee to apogee of the Moon's orbit, 239 anomalistic months. The distance will determine the magnitude of the eclipse, ensuring the similarity of eclipses at the period of the cycle. The Saros cycle is not an integer number of days ($6,585 \frac{1}{3}$), causing the eclipses in successive cycles to be displaced by eight hours in time (and solar eclipses, only visible at limited geographical locations, to be displaced by 120° in longitude). True repeats come after 3 Saros cycles, the 54-year Exeligmos cycle, but not with identical solar eclipse paths.

the key to understanding the function⁶ of the upper back dial. The references to “golden little sphere” and “little sphere” probably refer to the front zodiac display for the Sun and Moon—including phase for the latter.

The text near the lower back dial includes “Pharos” and “from south (about/around)....Spain (ΙΣΠΑΝΙΑ) ten”. These geographical references, together with previous readings¹ of “towards the east”, “west-north-west” and “west-south-west” suggest an eclipse function for the dial, as solar eclipses occur only at limited geographical sites, and winds were often recorded^{17–19} in antiquity with eclipse observations. Possibly this information was added to the mechanism during use.

Turning to the dials themselves, the front dial displays the position of the Sun and Moon in the zodiac, and a corresponding calendar¹ of 365 days that could be adjusted for leap years. Previously¹, it was suggested that the upper back dial might have five concentric rings with 47 divisions per turn, showing the 235 months of the 19-year Metonic cycle. A later proposal⁵ augments this with the upper subsidiary dial showing the 76-year Callippic cycle. Our optical and X-ray microfocus computed tomography (CT) imaging confirms these proposals, with 34 scale markings discovered on the upper back dial. On the basis of a statistical analysis analogous to that described for gear tooth counts below, we confirm the 235 total divisions. We also find from the CT that the subsidiary dial is indeed divided into quadrants^{1,6}, as required for a Callippic dial. In agreement with the back door inscription, we also substantiate the perceptive proposal^{5,20} that the dial is in fact a spiral, made from semicircular arcs displaced

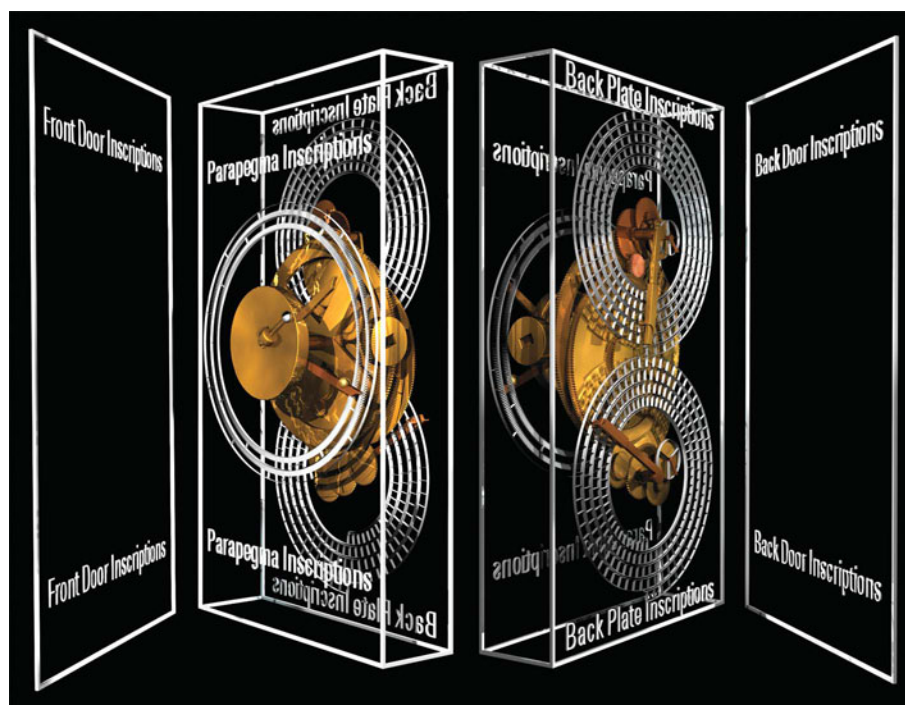


Figure 2 | A schematic view of the mechanism to illustrate the position of major inscriptions and dials. The front dial has two concentric scales. The inner scale shows the Greek zodiac with 360 divisions. There are occasional Greek letters denoting references to the Parapegma inscription, and we add three further reference letters (Z, H, Θ) to Price's description¹. The Parapegma is a star almanac showing rising and settings at dawn or evening of particular stars or constellations, which we will discuss elsewhere. Its form is consistent with a date of late second century BC. The outer (originally) movable scale is a calendar carrying the Egyptian names of the months with Greek letters. The Egyptian calendar of 365 days, with twelve 30-day months and 5 extra (epagomenai) days was in standard use in Greek astronomy. The effect of the extra quarter day in a year could be corrected by turning the

scale one day every four years—and a sequence of holes to take a locking pin is observed under the scale. We find that spacing of the holes is indeed what would be expected for a total of 365 days, with a possible range 363–365. The position of the Sun and Moon would have been indicated by pointers across the dial scales, and a device⁷ showing the phase of the Moon was probably carried round on the lunar pointer. It is not clear whether the Sun position pointer would have been separated from a date pointer, or whether any planetary positions might have been displayed. The spiral upper back dial displays the luni-solar Metonic sequence of 235 lunar months with a subsidiary dial showing the Callippic cycle, while the spiral lower back dial displays the 223-lunar-month Saros eclipse cycle with a subsidiary dial showing the Exeligmos cycle.

to two centres on the vertical midline. In the CT of fragment B we find a new feature that explains why the dial is a spiral: a 'pointer-follower' device (Fig. 3) travelled around the spiral groove to indicate which month (across the five turns of the scale) should be read.

From our CT data of the 48 scale divisions observed in fragments A, E and F, we establish 223 divisions in the four-turn^{5,20} spiral on the lower back dial, the spiral starting at the bottom of the dial. This is the Saros eclipse cycle, whose number is on the back door inscription. The 54-year Exeligmos cycle of three Saros cycles is shown on the lower subsidiary dial.

Between the scale divisions of the Saros dial we have identified 16 blocks of characters, or 'glyphs' (see Supplementary Note 2 ('glyphs and inscriptions')) at intervals of one, five and six months. These are eclipse predictions and contain either Σ for a lunar eclipse (from $\Sigma\epsilon\alpha\eta\eta\eta$, Moon) or H for a solar eclipse (from $\eta\alpha\iota\omicron\sigma$, Sun) or both. A correlation analysis (analogous to DNA sequence matching) with historic eclipse data²¹ (all modern eclipse data and predictions in our work are from this reference) indicates that over a period of 400–1 BC, the sequence of eclipses marked by the identified glyphs would be exactly matched by 121 possible start dates. The matching only occurs if the lunar month starts at first crescent, and confirms this choice of month start in the mechanism. The sequences of eclipses can then be used to predict the expected position of glyphs on the whole dial, as seen in Fig. 4. The dial starts and finishes with an eclipse. Although Ptolemy indicates that the Greeks recorded eclipses in the second century BC, the Babylonian Saros canon^{17–19} is the only known source of sufficient data to construct the dial.

The functions of the mechanism are determined by the tooth counts of the gears. These are based mainly on the CT, using angular measurement from a nominal centre to the remains of tooth tips. In a few cases all teeth can be seen, but many gears are incomplete. Counts are established by fitting models with regularly spaced teeth and minimising the r.m.s. deviation from the measurements—varying the centre in software (when unclear) to find the best-fit solution

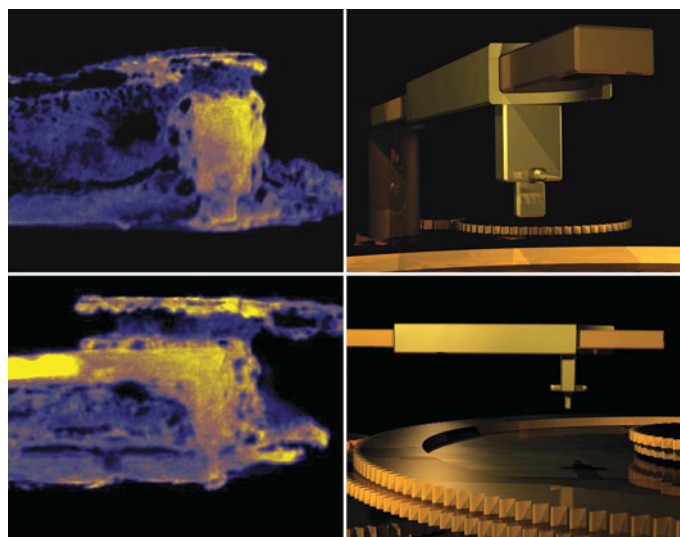


Figure 3 | The 'pointer-follower' lunar month indicator of the upper back dial. On the left, false-colour sections through CT images, analysed with VGStudio Max software by Volume Graphics GmbH. These show two views at right angles of the pointer-follower in the Metonic dial in fragment B. On the right, a computer reconstruction of the device from two different angles (with the Metonic scale omitted for clarity). The pin was constrained to follow the groove between the spiral scales (the scale is shown in Fig. 4), causing the device to slide along the month pointer to indicate which ring on the spiral scale specified the month. A similar pointer-follower would have been present on the lower back (Saros) dial. The Metonic dial would have required re-setting every 19 years, the Saros dial after 18 years. The groove-pin may have been held in place by the small pin through the front of the device, enabling its removal for re-setting.

or solutions (see Supplementary Note 3 ('gears')). We have adopted a systematic nomenclature of lower case letters for the axis of the gear, with numbering increasing with ordering from the front of the mechanism. Hypothetical (lost) gears are denoted by italics.

Several models have been proposed for the gear trains^{1,2,4–6,8}. We agree with the assumption of four missing gears ($n1$, $n2$, $p1$, $p2$) to drive the Metonic and Callippic dials⁴. We propose a new reconstruction for the other trains, which uses all extant gears (except the lone r1



Figure 4 | Reconstruction of the back dials. A composite of fragments A, B, E and F. The Metonic calendar is at top, with its subsidiary Callippic dial. The Saros eclipse cycle is below, with its subsidiary Exeligmos dial. The 16 observed eclipse glyphs are shown in turquoise on the Saros dial, with 35 hypothetical glyphs in violet. The hypothetical glyphs are based on the criterion that 99% of the 121 sequences exactly matching the observed glyphs have an eclipse at the month position. Both main dials would have a 'pointer-follower' (see Fig. 3) to indicate the relevant lunar month on the spiral. The monthly divisions on the Metonic upper back dial are not simply scribed directly across all five turns, as might be expected for simplicity of construction. There are small misalignments, implying a systematic attempt at marking full (30-day) and hollow (29-day) months. The incomplete data do not allow good analysis, other than a hint of bimodality in the interval distribution. If the marking out of the scale were carried out using the mechanism's gearing, then this would greatly pre-date known 'dividing engines'²⁷ by many centuries.

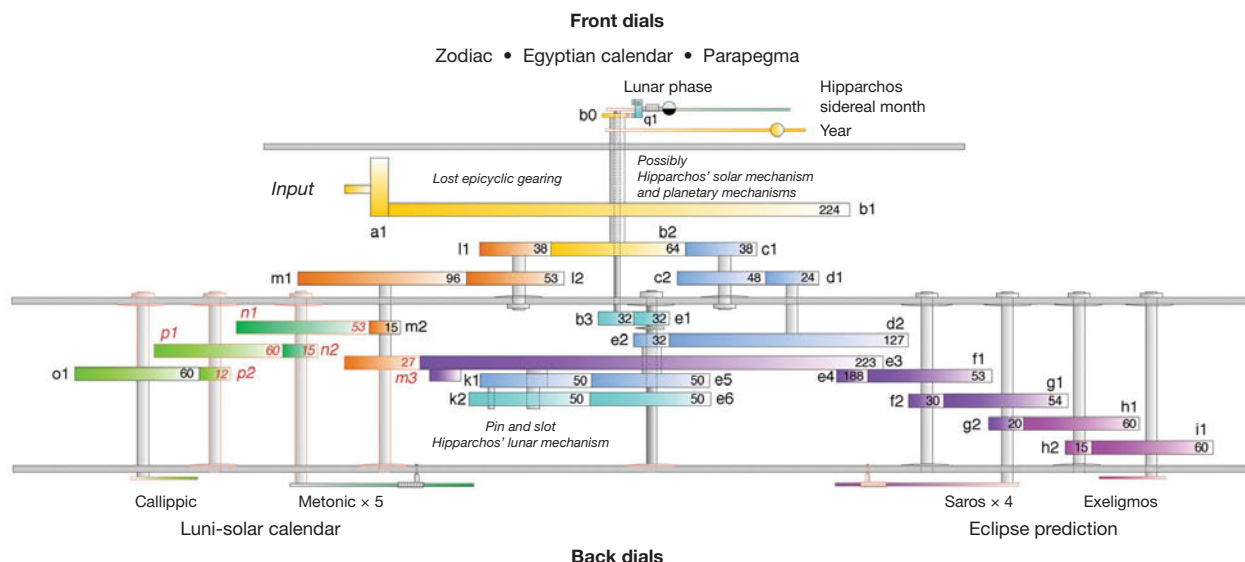


Figure 5 | New reconstruction of the gear trains. A schematic sectional diagram (not to scale) of the gearing, following the style of Price¹ and Wright⁴. The viewpoint is looking down from the top right of the mechanism, and is stretched in the direction of the main axes to show the structure. Features that are outlined or labelled in red are hypothetical. Gears are lettered with their shaft, and numbered with increasing distance from the front dial. The two-or-three digit number on the gear is its actual or assumed tooth count (see Supplementary Note 3 ('gears')). Hypothetical gears *n1*, *n2*, *p1*, *p2* have been proposed previously, the gear *m3* on the

from the separate fragment D). The proposed model is shown in Fig. 5. We require the assumption of only one further gear (*m3*), whose proposed shaft is clearly broken off in the CT. A detailed description is contained in Supplementary Note 3 ('gears').

Of particular note is the dual use of the large gear, *e3*, at the back of the mechanism, which has found no use in previous models. In our model, it is powered by *m3* as part of a fixed-axis train that turns the Saros and Exeligmos dials for eclipse prediction, and also doubles as the 'epicyclic table' for the gears *k1*, *k2*. These are part of epicyclic

broken-off shaft *m* is our addition. All gears, except the lone one in fragment D, are now accounted for in the mechanism. The function of the trains is outlined in the text. We find no evidence in the CT for an idler wheel carried on *e3* and between *e5* and *k1* or between *k2* and *e6*, as has been previously proposed^{1,2,4}. The CT shows a pin through axis *e* between gears *e1* and *e2*. We believe its purpose is to retain the square-bossed *e1* on the shaft, but its passage right through the axis rules out previous reconstructions^{1,2,4} where *e1* and *e2* were joined by an outer pipe rotating around the shaft *e*.

gearing that calculates the theory of the irregular motion of the moon, developed by Hipparchos some time between 146 and 128 BC (ref. 22)—the 'first anomaly', caused by its elliptical orbit about the Earth. The period of this anomaly is the period from apogee to apogee (the anomalistic month). To realize this theory, the mean sidereal lunar motion is first calculated by gears on axes *c*, *d* and *e* and this is then fed into the epicyclic system. As explained in Fig. 6, a pin-and-slot device on the epicyclic gears *k1* and *k2*, clearly seen in the CT, provides the variation. This was previously identified⁴, but rejected as a lunar mechanism. The remarkable purpose of mounting the pin-and-slot mechanism on the gear *e3* is to change the period of variation from sidereal month (that is, the time taken for the Moon to orbit the Earth relative to the zodiac), which would occur if *k1* and *k2* were on fixed axes, to anomalistic month—by carrying the gears epicyclically at a rate that is the difference between the rates of the

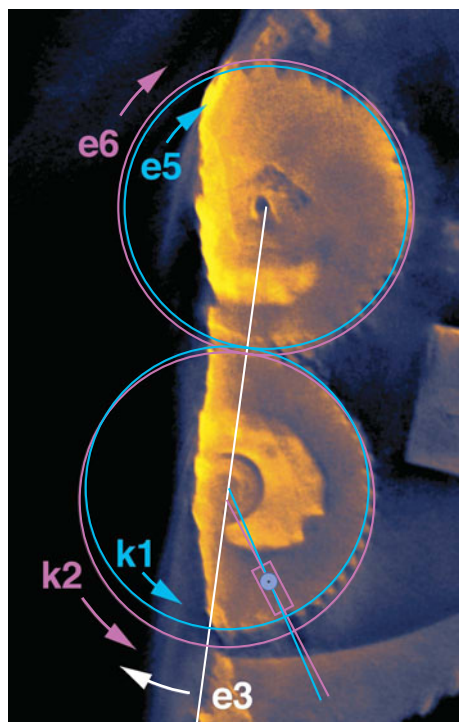


Figure 6 | The 'Hipparchos' lunar mechanism mounted on gear *e3*. The figure is based on a CT slice of part of fragment A, showing (top) shaft *e* and (bottom) shaft *k*. The complete geometry cannot be seen in a single CT slice. The two gears on the *e* axis (*e5* and *e6*) are coaxial, while the two *k* gears rotate on slightly displaced axes. *k1* has a pin on its face that engages with a radial slot in *k2* (and this was previously reported⁵). In the figure the pitch circles of *e5* and *k1* are shown in turquoise and those of *e6* and *k2* in pink. The gear *e5* drives *k1*, which drives *k2* via the pin-and-slot, introducing a quasi-sinusoidal variation in the motion, which is then transmitted to *e6*. Our estimate of the distance between the arbors on the *k* gears is about 1.1 mm, with a pin distance of 9.6 mm, giving an angular variation of 6.5°. According to Ptolemy²⁸, Hipparchos made two estimates for a lunar anomaly parameter, based on eclipse data, which would require angular variations of 5.9° or 4.5° here—although estimates of the anomaly from Babylonian astronomy were generally larger. The difference from our estimated value is probably not significant given the difficulty of precise measurement of the axes in the CT. The harmonic variation, together with the effect of carrying the gears on *e3* (which rotates at the period of the Moon's apogee around the Earth), would simulate the correct variation for the Moon's mean (sidereal) rotation rate on the front dial. An (unexplained) regular pentagon is visible at the centre of gear *e5*. It is tempting to associate the conception of the mechanism with Hipparchos himself, but he was not the first to assume eccentric or epicyclic models.

sidereal and anomalistic months, that is, at the rate of rotation of about 9 years of the Moon's apogee.

Gears with 53 teeth are awkward to divide. So it may seem surprising that the gearing includes two such gears (f1, l2), whose effects cancel in the train leading to the Saros dial. But the gearing has been specifically designed so that the 'epicyclic table' e3 turns at the rate of rotation of the Moon's apogee—the factor 53 being derived from the calculation of this rotation from the Metonic and Saros cycles, which are the bases for all the prime factors in the tooth counts of the gears. The establishment of the 53-tooth count of these gears is powerful confirmation of our proposed model of Hipparchos' lunar theory. The output of this complex system is carried from e6 back through e3 and thence, via e1 and b3, to the zodiac scale on the front dial and the lunar phase⁷ mechanism. Our CT confirms the complex structure of axis e that this model entails.

The Antikythera Mechanism shows great economy and ingenuity of design. It stands as a witness to the extraordinary technological potential of Ancient Greece, apparently lost within the Roman Empire.

Received 10 August; accepted 17 October 2006.

- Price, D. de S. Gears from the Greeks: The Antikythera Mechanism — A calendar computer from ca. 80 BC. *Trans Am. Phil. Soc. New Ser.* 64, 1–70 (1974); reprinted by Science History Publications, New York (1975).
- Wright, M. T. Epicyclic gearing and the Antikythera Mechanism, Part I. *Antiquar. Horol.* 27, 270–279 (2003).
- Wright, M. T., Bromley, A. G. & Magou, H. Simple X-ray tomography and the Antikythera Mechanism. *PACT J. Eur. Study Group Phys. Chem. Biol. Math. Tech. Appl. Archaeol.* 45, 531–543 (1995).
- Wright, M. T. The Antikythera Mechanism: a New Gearing Scheme. *Bull. Sci. Instrum. Soc.* 85, 2–7 (2005).
- Wright, M. T. Epicyclic gearing and the Antikythera Mechanism, Part II. *Antiquar. Horol.* 29, 51–63 (2005).
- Wright, M. T. Counting months and years: The upper back dial of the Antikythera Mechanism. *Bull. Sci. Instrum. Soc.* 87, 8–13 (2005).
- Wright, M. T. The Antikythera Mechanism and the early history of the moon-phase display. *Antiquar. Horol.* 29, 319–329 (2006).
- Wright, M. T. in *Proc. 2nd Int. Conf. on Ancient Greek Technology* (ed. Tassios, Th.) 49–60 (Technical Chamber of Greece, Athens, 2006).
- Wright, M. T. A planetarium display for the Antikythera Mechanism. *Horol. J.* 144(5), 169–173 (2002); erratum 144, 193 (2002).
- North, J. D. *The Fontana History of Astronomy and Cosmology* (Fontana, London, 1994).
- Bromley, A. G. The Antikythera Mechanism. *Horol. J.* 132, 412–415 (1990).
- Bromley, A. G. in *Bassernet Vol. 2, No. 3* (Basser Department of Computer Science, Univ. Sydney, June, 1993).
- Freeth, T. The Antikythera Mechanism: 1. Challenging the classic research. *Mediterr. Archaeol. Archaeom.* 2, 21–35 (2002).
- Edmunds, M. & Morgan, P. The Antikythera Mechanism: still a mystery of Greek astronomy? *Astron. Geophys.* 41, 6.10–6.17 (2000).
- Freeth, T. The Antikythera Mechanism: 2. Is it Posidonius' orrery? *Mediterr. Archaeol. Archaeom.* 2, 45–58 (2002).
- Illsley, J. S. History and Archaeology of the Ship—Lecture Notes. (<http://cma.soton.ac.uk/HistShip/shlect36.htm>) (1998).
- Stephenson, F. R. *Historical Eclipses and Earth's Rotation* (Cambridge Univ. Press, Cambridge, UK, 1997).
- Steele, J. M. *Observations and Predictions of Eclipse Times by Early Astronomers* (Kluwer Academic, Dordrecht, 2000).
- Steele, J. M. Eclipse prediction in Mesopotamia. *Arch. Hist. Exact Sci.* 54, 421–454 (2000).
- Wright, M. T. The Scholar, the Mechanic and the Antikythera Mechanism. *Bull. Sci. Instrum. Soc.* 80, 4–11 (2003).
- Esenak, F. NASA Eclipse Home Page. (<http://sunearth.gsfc.nasa.gov/eclipse/eclipse.html>) (2005).
- Jones, A. The adaptation of Babylonian methods in Greek numerical astronomy. *Isis* 82, 440–453 (1991).
- Britton, J. P. in *Die Rolle der Astronomie in den Kulturen Mesopotamiens* (ed. Galter, H. D.) 61–76 (rm-Druck & Vergangengesellschaft, Graz, 1993).
- X-Tek Systems Ltd. 3-D computed tomography. (<http://www.xtek.co.uk/ct/>) (2006).
- Malzbender, T. & Gelb, D. Polynomial texture mapping. (<http://www.hpl.hp.com/research/ptm/>) (2006).
- Brooks, M. Tricks of the light. *New Sci.* 170, 38–49 (2001).
- Chapman, A. *Dividing the Circle* (Wiley, Chichester, 1995).
- Toomer, G. J. *Ptolemy's Almagest* (transl. Toomer, G. J.) (Princeton Univ. Press, Princeton, New Jersey, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was financed by the Leverhulme Trust, the Walter Hudson Bequest, the University of Athens Research Committee and the Cultural Foundation of the National Bank of Greece. For essential support we thank the Ministry of Culture, Greece (P. Tatoulis), and the National Archaeological Museum of Athens (N. Kaltsas). We acknowledge help and advice from J. Ambers, J. Austin, G. Dermody, H. Forsyth, I. Freestone, P. Haycock, V. Horie, A. Jones, M. Jones, P. Kipourou, H. Kritzas, J. Lossi, G. Makris, A. Ray, C. Reinhart, A. Valassopoulos, R. Westgate, T. Whiteside, S. Wright and C. Xenikakis.

Author Contributions T.F. carried out most of the CT analysis of structure and its interpretation. Y.B., A.T. and X.M. read, transcribed and translated the inscriptions. H.M. and M.Z. catalogued the fragments, provided guidance on X-ray examination, and measured the fragments with J.H.S. R.H. led the team (D.B., A.R., M.A., A.C. and P.H.) that built and operated the Bladerunner CT machine, and provided CT reconstructions and advice. T.M., D.G. and W.A. built, operated and provided software for the PTM. M.G.E. was academic lead, and undertook the statistical analysis. T.F. and Y.B. organised the logistics of the experimental work, with inter-agency liaison by X.M. and J.H.S. The manuscript was written by T.F. and M.G.E. including material from Y.B., A.T., X.M., J.H.S., H.M. and M.Z. T.F. designed the illustrations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.G.E. (mge@astro.cf.ac.uk).

LETTERS

Rank clocks

Michael Batty¹

Many objects and events, such as cities, firms and internet hubs, scale with size^{1–4} in the upper tails of their distributions. Despite intense interest in using power laws to characterize such distributions, most analyses have been concerned with observations at a single instant of time, with little analysis of objects or events that change in size through time (notwithstanding some significant exceptions^{5–7}). It is now clear that the evident macro-stability in such distributions at different times can mask a volatile and often turbulent micro-dynamics, in which objects can change their position or rank-order rapidly while their aggregate distribution appears quite stable. Here I introduce a graphical representation termed the ‘rank clock’ to examine such dynamics for three distributions: the size of cities in the US from AD 1790, the UK from AD 1901 and the world from 430 BC. Our results destroy any notion that rank–size scaling is universal: at the micro-level, these clocks show cities and civilizations rising and falling in size at many times and on many scales. The conventional model explaining such scaling on the basis of growth by proportionate effect cannot replicate these micro-dynamics, suggesting that such models and explanations are considerably less general than has hitherto been assumed.

I begin with the US city size distributions compiled⁸ for the populations of the largest 100 cities from AD 1790 to AD 2000 using the decennial Population Census. City sizes are represented as rank–size distributions after Zipf, where population in city i at time t , $P_i(t)$, is ordered against its rank $r_i(t)$ and plotted logarithmically to give an immediate visualization of scaling⁹. The relative stability of these rankings is clear from the plots in Fig. 1a but the switch in rankings, with many cities entering and leaving the top 100, is completely hidden. Over this 210 yr period, there are 266 cities that at some stage belong to the top 100; from 1840 when the number of cities first reached 100, only 21 remain in 2000. On average, it takes 105 yr for 50% of cities to appear or disappear from the top 100, while the average change in rank order for a typical city in each 10 yr period is 7 ranks. If the distributions are collapsed onto one another, there is only an 18% difference between any of their rank orders. This contrasts with a typical switch in rank orders, which is also illustrated in Fig. 1a, where I plot the ‘1950’ ranks using ‘2000’ population values.

To visualize these micro-dynamics, I first plot trajectories in Fig. 1b based on the rank and size of each city in the rank–size space where each city is coloured according to its rank order and the time when it first appears in the top 100. These trajectories are ordered so that greater changes in rank overlay lower ones. Although cities that do not change their rank show up clearly as vertical lines, this plot is confusing. I therefore propose a ‘rank clock’, where rank orders are plotted for each city in temporal clockwise direction with the highest rank at the centre and the lowest on the circumference. This provides an immediate visualization of the dynamics, as I show for the US data in Fig. 1c. To develop meaningful analysis on the clock, I measure various displacements of rank, but first extract significant city trajectories as in Fig. 1d. New York City has been the top rank (at the clock’s centre) since 1790, but the clock reveals how large cities such as Chicago, Los Angeles and Houston enter the system as population

diffuses across the US, how cities in the original 13 colonies such as Richmond and Charleston fall out of favour, and how cities in the northeast ‘rust-belt’ such as Buffalo slowly decline in rank.

To complement the US data, I have constructed clocks for two other very different data sets. The US data are based on the rapid growth of key cities in the New World, and represent the transition from an essentially agrarian society to an industrial one. The original 24 cities in 1790 comprise only 5% of the US population (~200,000), whereas in 2000, the top 100 cities constitute 20% (~55 million). The UK data, taken from a reworking of the decennial Census data into 458 urban places¹⁰, illustrates a population growing from 37 million to 57 million over the period AD 1901 to AD 2001. The third set is based on Chandler’s 4,000 yr world history of urban growth¹¹, from which I have culled and added data to the top 50 cities from 430 BC to AD 2000. There are 390 distinct cities in this data that grow from about 3 million in total to almost 600 million over the period. Like the US data, this reveals the massive transition to industrialization but only at the very end of the time series. Other key transitions from the classical era to the Dark Ages, the importance of China and Japan, the ebb and flow of urban populations in the Middle Ages, and the emergence of the modern era, are all reflected in this data. Details are contained in the Supplementary Information.

Rank–size distributions and rank clocks are shown for the UK and world data in Fig. 2. The UK rank–size profiles show scaling in the upper tail as the number of urban places in this data set is fixed. In the UK data, of the top 100 cities in 1901, 73 remain in these top ranks in 2001. The change in rank order in each 10 yr period for an average city is only 14 out of a total of 458 cities, which is less than half the US rate. In the world data, no cities in the top 50 in 430 BC remain in the top 50 in the year 2000, while from the Fall of Constantinople in 1453, only 6 cities from that era now appear in the top 50 (in 2000). The cities that show the greatest longevity through their being in the top 50 over the 2,430 yr period are perhaps surprising: Suzhou (2,158 yr), Nanking (2,080 yr), Wuchang (1,850 yr), Benares (1,780 yr), Rayy (1,630 yr) and Rome (1,530 yr). Paris (525 yr) is 77th out of 390 cities.

The half life for cities in this data set remaining in the top 50 is ~200 yr, almost twice as long as for the US data. Comparing the patterns in the three clocks in Figs 1 and 2, visual intuition suggests that changes in rank are smoothest in the UK data, more volatile and complex in the US, while most volatile in the world data where large changes in rank seem to dominate all historical periods. A more detailed sample of this micro-dynamics is shown in the clock in Fig. 3, where 7 ‘typical’ cities are plotted for the world data, each showing a different pattern of rank change, suggesting a possible basis for future classification of city types based on their local dynamics¹².

To compare these three systems, I reduce each set of cities to the top 50. These of course generate different rank clocks (shown in the Supplementary Information) in the US and UK systems, as I am now dealing with the top 50 cities from 266 US and 458 UK cities. Throughout I am also making the assumption that a city that exists at times $t-1$ and t in the top 50 remains in that list regardless of the length of time between $t-1$ and t . For the world data set where time

¹Centre for Advanced Spatial Analysis, The Bartlett School, University College London, 1–19 Torrington Place, London WC1E 6BT, UK.

periods vary substantially in length, this complicates the analysis, for cities might hop in and out of the top 50 during long time intervals.

The rank shifts¹³ are defined as distances $d_i(t) = |r_i(t) - r_i(t-1)|$, which exist only if the city in question is in the top 50 at each time period within which $N_i(t)$ such cities comprise the common set $\Omega(t)$. These distances can be plotted on the clock, as can the average shift per time $d(t) = \sum_{i \in \Omega(t)} |r_i(t) - r_i(t-1)| / N_i(t)$, which enables analysis of the extent to which the system is converging or diverging in terms of rank shift¹⁴. The average shift over all time periods T is defined as $d = \sum_t d(t) / T$. If there are no shifts, then all these distances are zero, while the maximum average shift that can take place at any time is $d(t) < n/2$ where $n = 50$, the total number of cities in the list.

These distances are plotted in Fig. 4a–c. The world system is clearly the most volatile, with an overall shift d of 14.27 and with the average $d(t)$ indicating bigger shifts in the classical era and until around the year AD 1000. After that, there is a gradual reduction in shift until the Industrial Revolution. For the US data, the overall shift is much lower with d at 4.67, with the average $d(t)$ being largest between 1830 and 1890, the period of greatest city expansion in the US. I would expect the UK to have much lesser volatility from the earlier analysis, but once the city set is restricted to the top 50, changes in rank (d at 4.22) are only a little less than for the US case. The UK was effectively locked into its current urban settlement pattern by 1901, and rank shifts in the last century are greatest in the 1950s and 1960s, a period of rapid suburbanization¹⁵.

My second analysis involves defining growth rates and examining their trajectories using the rank clock. These rates will be split into

those associated with the growth and the change in the share of population in cities. First I define the growth rate for each city i at time t , $\lambda_i(t)$, as $P_i(t)/P_i(t-1)$, from which $P_i(t) = \lambda_i(t) P_i(t-1)$ and then define population shares as $p_i(t) = P_i(t)/P(t)$ where total population is $P(t) = \sum_i P_i(t)$. The expected growth rate for cities $\lambda(t)$ is:

$$\lambda(t) = \sum_i p_i(t) \lambda_i(t) = \sum_i p_i(t) \frac{P_i(t)}{P_i(t-1)} \quad (1)$$

$$= \left\{ \frac{P(t)}{P(t-1)} \right\} \left\{ \sum_i p_i(t) \frac{p_i(t)}{p_i(t-1)} \right\}$$

where the overall growth $\Gamma(t)$ and expected shift in population shares $\mathcal{G}(t)$ are the first and second terms on the second line of equation (1), respectively. A more workable form is based on the logarithmic growth rate $\log \lambda_i(t) = \log[P_i(t)/P_i(t-1)]$, whose expected value is an information statistic defined as:

$$I[\lambda(t)] = \sum_i p_i(t) \log \lambda_i(t) = \sum_i p_i(t) \log \frac{P_i(t)}{P_i(t-1)} \quad (2)$$

$$= \log \frac{P(t)}{P(t-1)} + \sum_i p_i(t) \log \frac{p_i(t)}{p_i(t-1)}$$

The log of overall growth $I[\Gamma(t)]$ and expected log of the shares $I[\mathcal{G}(t)]$ are the first and second terms on the second line of equation (2). The second term is an information difference, which holds the key to further analysis through decomposition into growth and change at different spatial scales, thereby enabling different systems to be integrated through a hierarchy of information entropies^{16–18}.

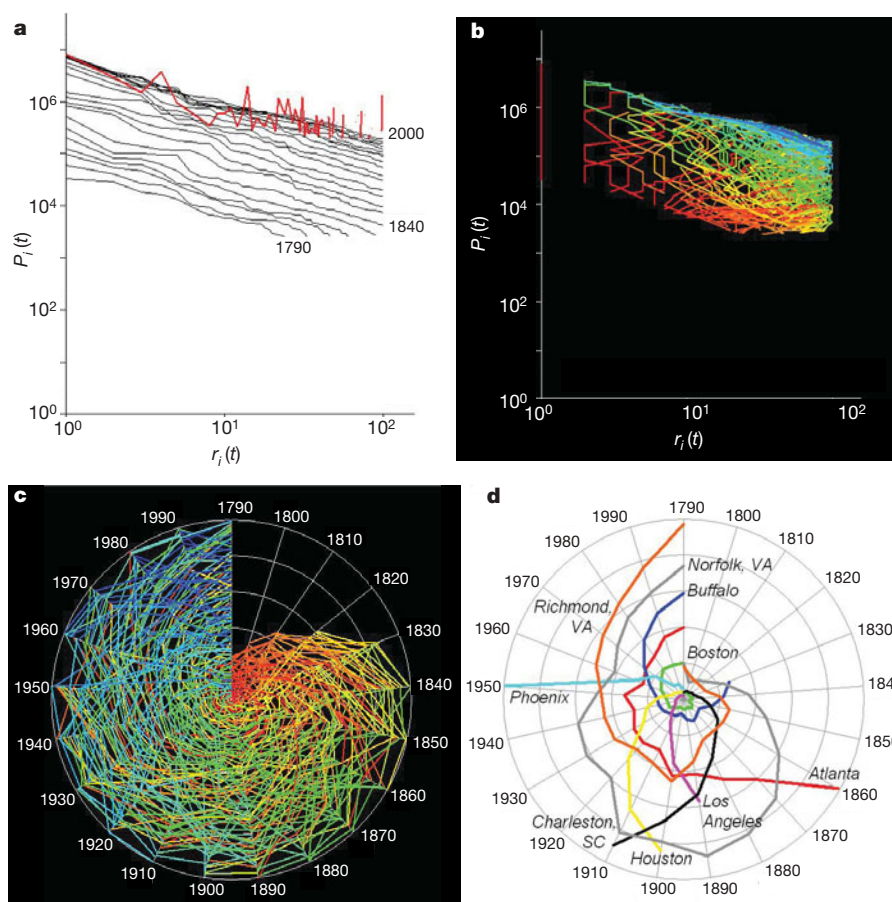


Figure 1 | Rank size and rank clocks for the US urban system 1790–2000. **a**, Zipf plots of the top 100 cities. Red indicates the rank of cities in 1790 using populations for 2000. **b**, Trajectories of each city in rank–size space with New York City rank 1 throughout. **c**, The rank clock, with each axis

running from rank 1 at the centre to 100 on the circumference, and cities coloured by date of entry from 1790 (red) to 2000 (blue). **d**, Sample city trajectories.

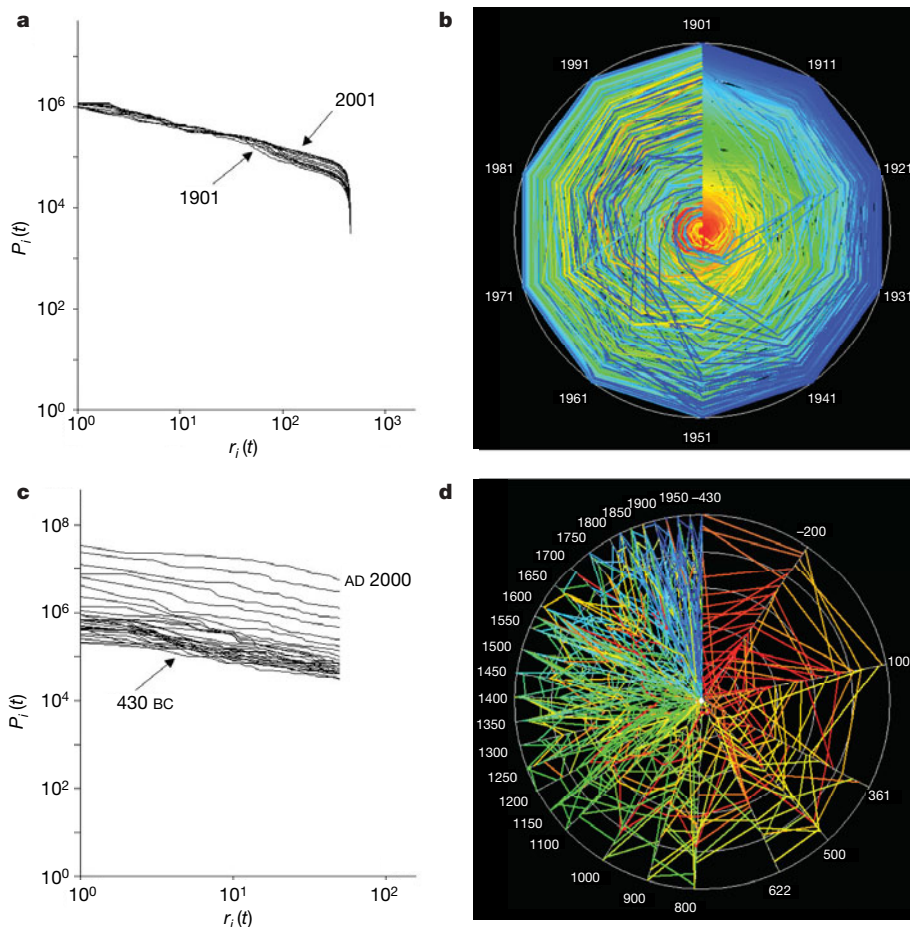


Figure 2 | Rank size and rank clocks for the UK and world urban systems. **a, b,** The Zipf plot (**a**) and the rank clock (**b**) for the UK, with each clock axis

from rank 1 to 458. **c, d,** The Zipf plot (**c**) and the rank clock (**d**) for the world, with each clock axis from rank 1 to 50.

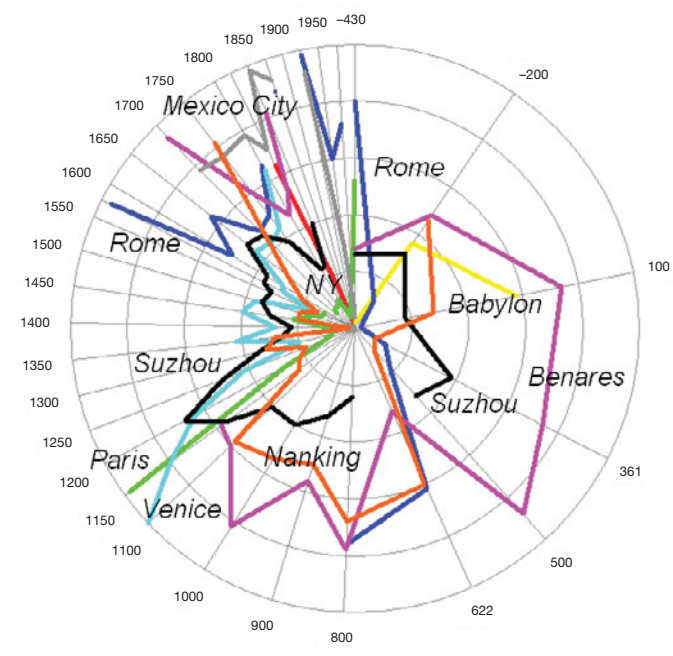


Figure 3 | Trajectories of a sample of world cities from 430 BC to AD 2000. Note the dominance of Rome until modern times, and the importance of the Chinese cities Nanking and Suzhou. From 1600 onwards, the volatility of the clock increases as cities of the early modern period and the Industrial Revolution are replaced in the top 50 by cities in the developing world.

These growth and information statistics reflect the balance between overall change and the shift and share in rank and population size, which in its simplest additive form is expressed in the expected log of city growth rates in equation (2). The growth and information components averaged over all time periods for each data set are shown in Table 1, where it is clear that for the world and UK systems, the averages $\mathcal{G}$ and $I[\mathcal{G}]$ are very small, implying relatively little contribution to overall growth, which mainly comes from Γ and $I[\Gamma]$. In contrast, there is greater shift and overall growth in the US system. However, the UK has been a low growth city system over the entire data period whereas the world system had relatively low growth from 430 BC to about AD 1600 when it began to take off. These patterns are very clear in the trajectories of the three growth components, which are plotted in rank clock form for all three systems in Fig. 4d. The information statistics mirror these trajectories. There we observe for the world system that the $\mathcal{G}(t)$ and $I[\mathcal{G}(t)]$ make very little contribution to the total change while $\Gamma(t)$ and $I[\Gamma(t)]$ dominate, with the significant periods being the early Dark Ages from the collapse of the Roman Empire to around 500, and the Industrial Revolution which merges into the current era of globalization beginning around 1800¹⁹. In the UK system, the statistics pick up the recession period 1930–40 and the relatively low growth era of the 1970s. Lastly, the US urban system shows much greater growth and population shifts in the nineteenth century with specific shifts in population in the 1860s and relatively low periods of growth in the 1820s, the 1930s and 1940s, and the 1980s.

The information generated on the micro-dynamics of rank size can be used to examine the consistency of the widely accepted model used to generate growth and change in systems that scale^{1,20–22}. This model is based on proportionate random growth, first suggested by Gibrat²³,

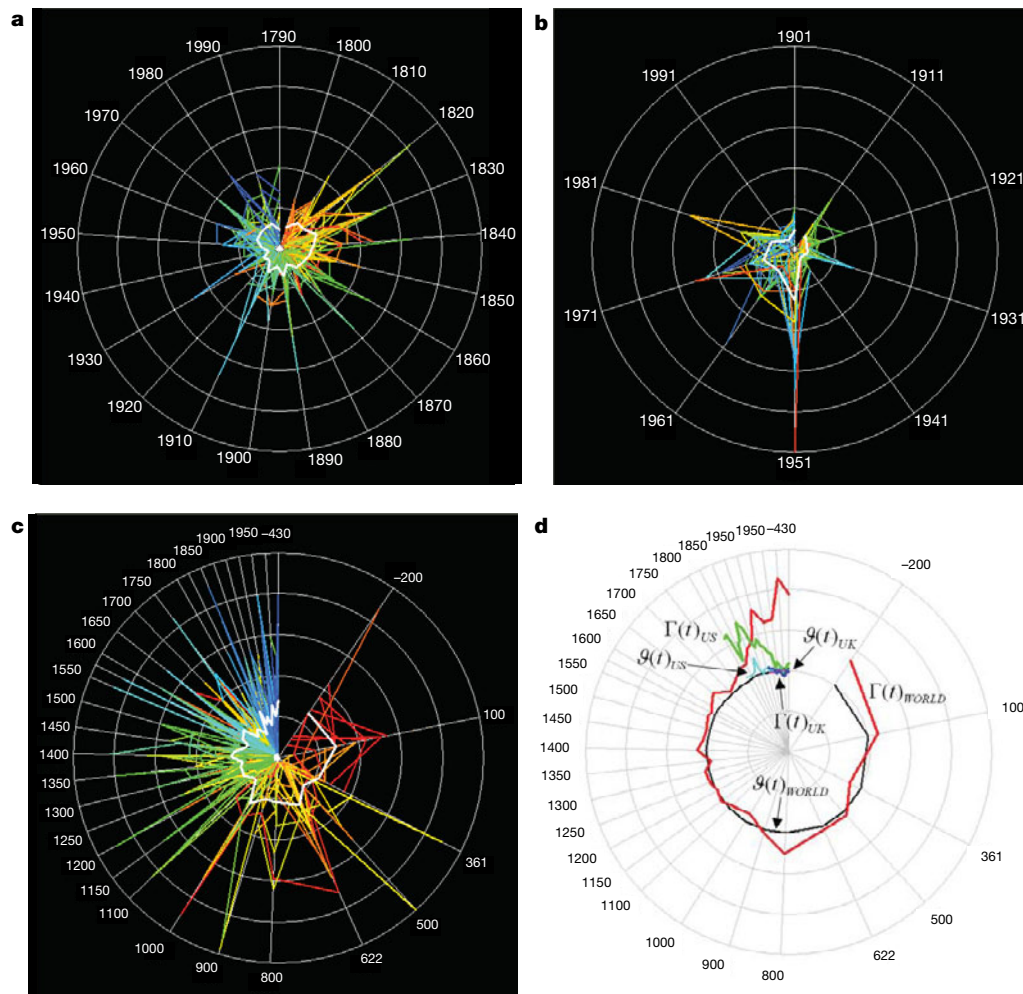


Figure 4 | Distance clocks and growth rates. **a–c**, Distance clocks for the top 50 cities for the US (**a**), the UK (**b**) and the world (**c**), coloured according to when each city enters the rank order from the earliest in red to the latest in

blue, and with greatest changes in distance overlaying lesser. **d**, The shift-share growth clock. All axes are from rank 1 to 50 for the clocks in **a**, **b** and **c**, and from 0 (centre) to 2.5 (circumference) for **d**.

subject to a lower bound $P_{\min}(t)$ below which populations cannot fall. For n cities, $P_i(t) = [I + \varepsilon_i(t)] P_i(t-1)$ with $P_i(t) > P_{\min}(t)$ where $\varepsilon_i(t)$ is a random value chosen from a normal distribution, and $P_{\min}(t)$ is a fixed proportion of the average $P(t)/n$ of the total population²⁴. With 1,500 cities and using a constant ten year growth rate of 1.126 discounted to each yearly time step, I have run the model for $T=2,500$ steps, which mirrors the population change in

Chandler's world data set. The biggest differences between the simulated and three real data sets are the low values of the population shift-shares with $\vartheta=1.00061$ and $I[\vartheta]=1.00031$. The macro- and micro-dynamics of this simulation are shown in the Supplementary Information, from which it is clear that the rank clock has some similarities with that associated with the world data set, whereas the distance clock is similar to the US clock.

It is clear however that this model is not able to generate the unique events associated with a turbulent world history in the rise and fall of cities and civilizations¹⁹, nor is it able to mirror change over shorter time periods. Although Gibrat's model does generate universal scaling behaviour for city size distributions, its micro-dynamics is very different from that which is revealed using the rank clocks. Models of proportionate random growth generating scale-free effects are now being explored for networks, and preliminary analysis of their historical dynamics can be informed using rank clocks²⁵. To test these ideas further, much larger temporal data sets—such as clusters of internet hubs and websites, firm sizes, scientific citation networks, individual income distributions, and short lived epidemics—would thus appear to be excellent candidates for analysis using rank clocks.

Received 12 May; accepted 27 September 2006.

- Blank, A. & Solomon, S. Power laws in cities population, financial markets and internet sites: Scaling and systems with a variable number of components. *Physica A* **287**, 279–288 (2000).
- Gabaix, X. & Ioannides, Y. M. in *Handbook of Regional and Urban Economics* Vol. 4 (eds Henderson, V. & Thisse, J.-F.) 2341–2378 (North-Holland, Amsterdam, 2004).

Table 1 | Distance, growth and information statistics

Component	US	World	UK	Gibrat's model
d	4.6675	14.2773	4.22049	9.4853
$\lambda = \sum_i \lambda_i(t) / T$	1.3787	1.1311	1.0044	1.3441
$\Gamma = \sum_i \Gamma_i(t) / T$	1.3122	1.1256	0.9991	1.3433
$\vartheta = \sum_i \vartheta_i(t) / T$	1.0456	1.0042	1.0054	1.0006
$I[\lambda] = \sum_i I[\lambda_i(t)] / T$	0.2770	0.1032	0.0010	0.2942
$I[\Gamma] = \sum_i I[\Gamma_i(t)] / T$	0.2572	0.1012	-0.0017	0.2939
$I[\vartheta] = \sum_i I[\vartheta_i(t)] / T$	0.0198	0.0020	0.0027	0.0003

See text for details of quantities in the leftmost column. All rates are normalized to 10 yr time intervals.

3. Axtell, R. L. Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820 (2001).
4. Adamic, L. A. & Huberman, B. A. Zipf's law and the internet. *Glottometrics* **3**, 143–150 (2002).
5. Stanley, M. H. R. *et al.* Scaling behavior in the growth of companies. *Nature* **379**, 804–806 (1996).
6. White, D. R., Kejzar, N., Tsallis, C. & Rozenblat, C. *Generative Historical Model of City-Size Hierarchies: 430BCE–2005* (ISCOM Working Paper, Institute of Mathematical Behavioral Sciences, University of California, Irvine, CA, 2005).
7. Barabasi, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
8. Gibson, C. *Population of the 100 Largest Cities and Other Urban Places in the United States: 1790 to 1990* (Population Division Paper 27, US Bureau of the Census, Washington DC, 1998).
9. Zipf, G. K. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, Massachusetts, 1949).
10. CDU census website. (<http://census.ac.uk/cdu/>) (accessed 8 August, 2006).
11. Chandler, T. *Four Thousand Years of Urban Growth: An Historical Census* (Edward Mellon, Lampeter, UK, 1987).
12. Ioannides, Y. M. & Overman, H. G. Zipf's law for cities: An empirical examination. *Reg. Sci. Urban Econ.* **33**, 127–137 (2003).
13. Havlin, S. The distance between Zipf plots. *Physica A* **216**, 148–150 (1995).
14. Guerin-Pace, F. Rank-size distribution and the process of urban growth. *Urban Stud.* **32**, 551–562 (1995).
15. Robson, B. T. *Urban Growth: An Approach* (Methuen, London, 1972).
16. Theil, H. *Statistical Decomposition Analysis* (North-Holland, Amsterdam, 1972).
17. Batty, M. Entropy in spatial aggregation. *Geogr. Anal.* **8**, 1–21 (1976).
18. Gell-Mann, M. & Tsallis, C. (eds) *Nonextensive Entropy-Interdisciplinary Applications* (Oxford Univ. Press, New York, 2004).
19. Turchin, P. *Historical Dynamics: Why States Rise and Fall* (Princeton Univ. Press, Princeton, New Jersey, 2003).
20. Gabaix, X. Zipf's law for cities: An explanation. *Q. J. Econ.* **114**, 739–767 (1999).
21. Sornette, D. & Cont, R. Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *J. Phys. I (Paris)* **7**, 431–444 (1997).
22. Pumain, D. in *Hierarchy in Natural and Social Sciences* (ed. Pumain, D.) 169–222 (Springer, Dordrecht, 2006).
23. Gibrat, R. *Les Inégalités Économiques* (Librairie du Recueil, Sirey, Paris, 1931).
24. Malcai, O., Biham, O. & Solomon, S. Power-law distributions and Levy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements. *Phys. Rev. E* **60**, 1299–1303 (1999).
25. Krapivsky, P. L. & Redner, S. Statistics of changes in lead node in connectivity-driven networks. *Phys. Rev. Lett.* **89**, 258703 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was partially supported by the EPSRC Spatially Embedded Complex Systems Engineering Consortium. I thank R. Carvalho for useful discussions, and D. Dorling for providing the UK data set.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (m.batty@ucl.ac.uk).

Active terahertz metamaterial devices

Hou-Tong Chen^{1*}, Willie J. Padilla^{1*†}, Joshua M. O. Zide², Arthur C. Gossard², Antoinette J. Taylor¹
& Richard D. Averitt^{1†}

The development of artificially structured electromagnetic materials, termed metamaterials, has led to the realization of phenomena that cannot be obtained with natural materials¹. This is especially important for the technologically relevant terahertz (1 THz = 10^{12} Hz) frequency regime; many materials inherently do not respond to THz radiation, and the tools that are necessary to construct devices operating within this range—sources, lenses, switches, modulators and detectors—largely do not exist. Considerable efforts are underway to fill this ‘THz gap’ in view of the useful potential applications of THz radiation^{2–7}. Moderate progress has been made in THz generation and detection⁸; THz quantum cascade lasers are a recent example⁹. However, techniques to control and manipulate THz waves are lagging behind. Here we demonstrate an active metamaterial device capable of efficient real-time control and manipulation of THz radiation. The device consists of an array of gold electric resonator elements (the metamaterial) fabricated on a semiconductor substrate. The metamaterial array and substrate together effectively form a Schottky diode, which enables modulation of THz transmission by 50 per cent, an order of magnitude improvement over existing devices¹⁰.

A great deal of research into metamaterials has used microwave radiation; this is in part due to the ease of fabrication of sub-wavelength structures at these frequencies. Indeed, negative refractive

index media^{11,12} composed of negative permittivity¹³ ($\epsilon_1 < 0$) and negative permeability¹⁴ ($\mu_1 < 0$) metamaterial elements was first demonstrated at microwave frequencies. This has led to intense theoretical, computational and experimental studies of exotic phenomena, such as perfect lensing¹⁵ and cloaking^{16,17}. Recently, researchers have ventured to create functional metamaterials at near-infrared and visible frequencies^{18–20}. Considerably less work has concentrated on THz frequencies^{21,22}. However, the design flexibility associated with metamaterials provides a promising approach, from a device perspective, towards filling the THz gap.

Metamaterials are geometrically scalable, which translates to operability over many decades of frequency. This engineering tunability is in fact a distinguishing and advantageous property of these materials. However, for many applications it is desirable to have real-time tunability. For instance, short-range wireless THz communication or ultrafast THz interconnects^{23,24} require switches and modulators. Current state-of-the-art THz modulators based on semiconducting structures have the desirable property of being broadband, which is of relevance to THz interconnects, but are only able to modulate a few per cent¹⁰ and usually require cryogenic temperatures²⁵. Therefore, further improvement of the performance characteristics are required for practical applications. Here we present an efficient active metamaterial switch/modulator operating at THz frequencies. Although the modulation is based on a narrowband metamaterial resonance,

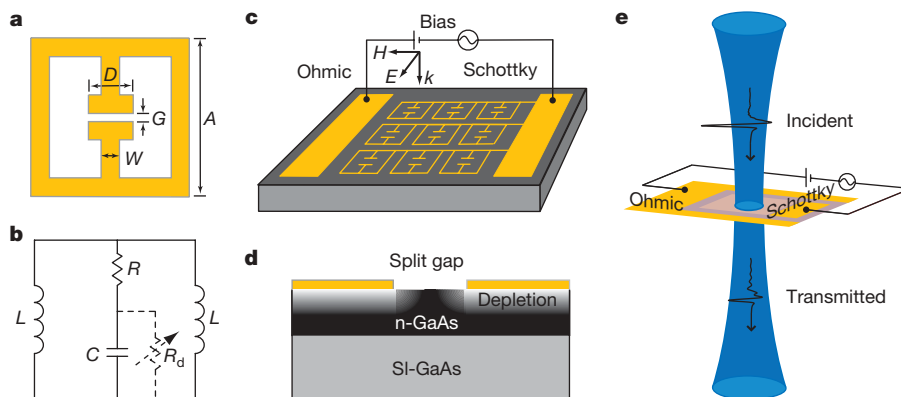


Figure 1 | Experimental design of the active THz metamaterial device.

a, Geometry and dimensions of the THz metamaterial switch/modulator: $A = 36 \mu\text{m}$, $G = 2 \mu\text{m}$, $D = 10 \mu\text{m}$ and $W = 4 \mu\text{m}$. **b**, An equivalent circuit of the metamaterial element, where the dashed variable resistor corresponds to loss due to the substrate free carrier absorption within the split gap. **c**, The metamaterial elements are patterned with a period of $50 \mu\text{m}$ to form a planar array of $5 \times 5 \text{ mm}^2$. These elements are connected together with metal wires to serve as a metallic (Schottky) gate. A voltage bias applied between the Schottky and ohmic contacts controls the substrate charge carrier density

near the split gaps, tuning the strength of the resonance. Orientation of the incident THz wave is indicated and the polarization of the electric field, E , magnetic field, H , and wave vector, k , are shown. **d**, Diagram of the substrate and the depletion region near the split gap, where the grey scale indicates the free charge carrier density. **e**, Experimental configuration for THz transmission measurements through a fabricated device. The black curves show the measured time-domain waveforms of the incident and transmitted THz pulses when a reverse gate voltage bias of 16 V is applied to the device and the THz electric field is polarized perpendicular to the connecting wires.

¹Center for Integrated Nanotechnologies, Materials Physics & Applications Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ²Materials Department, University of California, Santa Barbara, California 93106, USA. [†]Present addresses: Department of Physics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, Massachusetts 02467, USA (W.J.P.); Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA (R.D.A.).

*These authors contributed equally to this work.

these devices can be engineered to operate at specific frequencies. This would enable, as an example, amplitude modulation of narrow-band devices, such as THz quantum cascade lasers, enabling near-term practical applications.

The metamaterial device used in this work is based on a recently presented electric analogue to split-ring resonators (SRRs)²⁶. The geometry and dimensions are shown in Fig. 1a. The element consists of two single SRRs put together on the split gap side. These two rings provide inductances, L , and the split gap provides a capacitance, C , which are depicted as an equivalent circuit in Fig. 1b. A frequency-dependent dielectric resonant response results when it is patterned on a suitable substrate to form a planar periodic array of subwavelength structures. The two inductive loops are oppositely wound and thus any magnetic response is cancelled, resulting in a net electric response. The resistor R models the dissipation in the gold split rings, and the variable resistor R_d (shown dashed) models dissipation due to the substrate free carrier absorption within the split gap²⁷.

In our device, the metamaterial elements are electrically connected using conducting wires such that the entire metamaterial array functions as a voltage gate, schematically depicted in Fig. 1c. This structure has been designed to enable voltage control of the conductivity of the substrate at the split gaps, thereby controlling the THz transmission. The substrate consists of a 1- μm -thick n-type gallium arsenide (GaAs) layer with a free electron density of $n = 1.9 \times 10^{16} \text{ cm}^{-3}$ grown on a semi-insulating gallium arsenide (SI-GaAs) wafer by molecular beam epitaxy (MBE), as detailed in Fig. 1d. The ohmic contact is fabricated by electron-beam deposition of 20 nm of nickel, 20 nm of germanium, and 150 nm of gold in sequence, followed by rapid thermal annealing at 350 °C for 1 min in a nitrogen atmosphere. Next, the planar electric resonator array is fabricated using conventional photolithography and electron-beam deposition of a 10-nm-thick adhesion layer of titanium on the GaAs substrate, followed by 200 nm of gold. The metal and n-GaAs form a Schottky junction and the connected metamaterial resonators serve as a metallic gate. Current–voltage (I – V) measurements confirm the Schottky character of the device (Supplementary Fig. 1).

Terahertz time-domain spectroscopy (THz-TDS)²⁸ was used to characterize the performance of the metamaterial device, and has been described elsewhere in detail²⁹. In our photoconductive THz-TDS experiment, a polyethylene lens focuses the linearly polarized THz beam onto the metamaterial sample to a diameter of about 3 mm, and a second polyethylene lens recollimates the transmitted THz beam, which is directed to a photoconductive receiver. The experiments were performed at room temperature in a dry air atmosphere. In THz-TDS, the time-varying electric field of the impulsive THz radiation is recorded, and the electric field spectral amplitude and phase are directly obtained by performing Fourier analysis. Measurements of the metamaterial device with respect to a suitable reference, as illustrated in Fig. 1e, allow determination of the complex transmission as a function of frequency, $\tilde{t}(\omega)$. Inversion of $\tilde{t}(\omega)$ further permits model-independent calculation of the frequency-dependent complex permittivity²⁸, $\tilde{\epsilon}(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$, where ϵ_1 and ϵ_2 are the real and imaginary portions, respectively.

All experiments were performed at normal incidence, with the THz magnetic field lying completely in-plane. The polarization of the THz electric field was either perpendicular or parallel to the split gaps (and connecting wires). The wires connecting the individual electric resonators are necessary (as described above) to provide electrical connectivity to the gate. Importantly, these connecting wires have little effect on the electromagnetic properties of the electric resonators when the THz electric field is normal to the connecting wires. This was confirmed by finite element simulations using commercial software, as shown in Fig. 2a and b. The electric field is strongly concentrated at the split gaps, and there is no significant surface current flowing along the connecting metal wires between electric resonators at the resonant frequency ($\sim 0.72 \text{ THz}$).

Without an applied gate bias, the device is not expected to display resonant behaviour associated with the electric resonators because the substrate free charge carriers short out the capacitive response associated with the gaps. Upon application of a voltage, a resonant transmission should result as carriers in the substrate are displaced from the gaps. The blue curve in Fig. 2c shows the frequency-dependent transmitted intensity at a reverse gate bias of 16 V, where the polarization of the incident THz electric field is perpendicular to the connecting wires. Two distinct resonances are observed. The 0.72 THz resonance is the LC response associated with circulating currents in each metamaterial element, while the resonance at 1.65 THz is due to in-phase dipolar currents along the 36 μm lengths of the elements²². The spectrum is consistent with that from the same structure fabricated on an SI-GaAs substrate with no free carriers (red curve) and with simulation (black dashed curve), as shown in Fig. 2c. In Fig. 2d the real permittivity $\epsilon_1(\omega)$ of the THz metamaterial devices is shown as extracted from the experimental data of Fig. 2c assuming a cubic unit cell²⁶.

The resonances are strongly dependent on gate bias, as shown in Fig. 3. With zero applied voltage to the gate (black curves), the metamaterial response is very weak and does not show significant

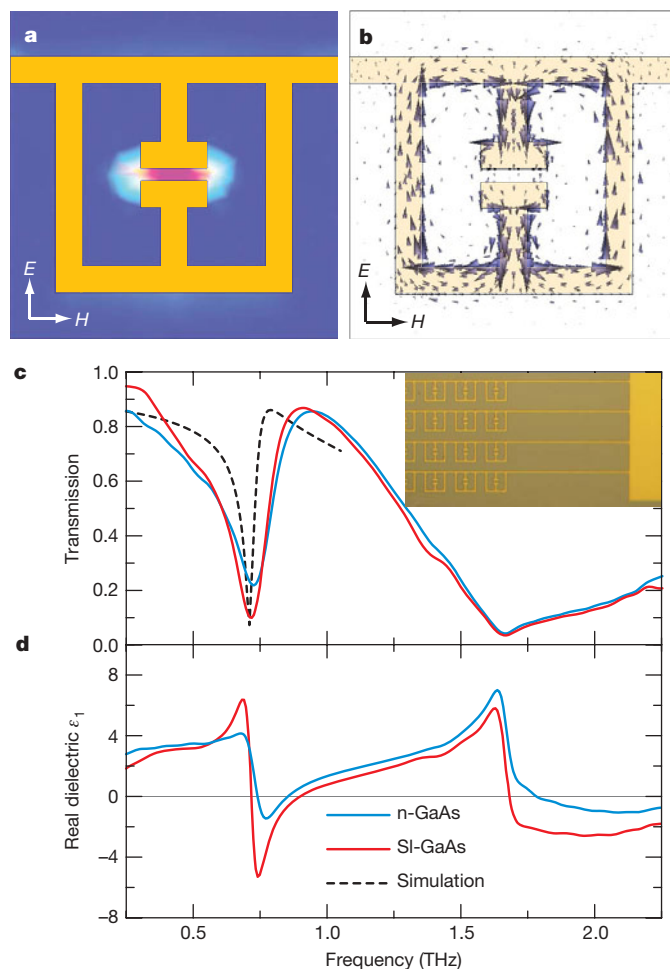


Figure 2 | Simulated and experimental characterization of THz metamaterial devices. **a**, Simulated norm of the electric field, $|E|$, and **b**, the surface current density of a metamaterial element at the 0.72 THz resonance. **c**, Frequency dependent transmitted intensity of THz radiation, and **d**, the corresponding extracted real part of the effective permittivity of the voltage controlled metamaterial device at 16 V reverse gate bias (blue curves) and the same structure fabricated on an SI-GaAs substrate (red curves). The dashed black curve is the simulated transmission for a device on the SI-GaAs substrate. The inset to **c** shows a photograph of the individual resonant elements, connecting wires and the contact pad, which together form the Schottky gate of the metamaterial device.

frequency dependence near the 0.72 THz resonance, as the relatively conductive substrate shorts the capacitive split gap and no LC resonance can be established. An increasing reverse gate bias depletes an increasing fraction of electrons in the n-GaAs layer near the metallic gate, thus significantly reducing the conductivity of the substrate near the split gaps, thereby restoring the LC resonant response. This is verified by the experimental results shown in Fig. 3a, as the resonances in the transmission spectra narrow and increase in amplitude with increasing reverse bias. At a reverse gate bias of 16 V, a 50% relative intensity change of transmission is observed at 0.72 THz, making this device a reasonably efficient narrowband THz switch/modulator. We note that the transmitted intensity of the 1.65 THz resonance also decreases with bias. This is because the substrate carriers screen the dipolar currents and is not associated with shorting of the capacitive gap of the metamaterial elements. Figure 3b shows the corresponding permittivity $\epsilon_1(\omega)$ of the metamaterial device at various gate biases. Clearly, $\epsilon_1(\omega)$ of the device is significantly modified by the applied gate bias. $\epsilon_1(\omega)$ increases on the low frequency side of the resonance while at higher frequencies it decreases to less than unity and even becomes negative.

From Fig. 3a, it is evident that at frequencies (~ 1 THz) between the two resonances, the transmission is significantly enhanced as a function of reverse gate bias. We have investigated whether this enhancement arises from a reduction of free carrier absorption in the n-GaAs layer due to depletion. For this purpose, we fabricated a device without the metamaterial array—only the connecting wires

remained as the Schottky contact. With the same polarization of the THz electric field, that is, perpendicular to the connecting wires as indicated by the inset to Fig. 3c, we measured its THz transmission at various bias voltages. As shown in Fig. 3c for reverse biases of 0 and 16 V, the change of THz transmission is hardly observable and variations are within the experimental noise. Additionally, we performed measurements of the free carrier absorption in the n-GaAs layer using an unpatterned sample with Si-GaAs as the reference. The relative intensity change of transmission between samples with carrier density $n = 1.9 \times 10^{16} \text{ cm}^{-3}$ and $n \approx 0$ is less than 10% at ~ 1 THz. Furthermore, in our metamaterial device only a small fraction of the n-GaAs layer is depleted by the reverse gate bias and thus the reduction of free carrier absorption is not nearly enough to account for the change in transmission in Fig. 3a. The transmission enhancement in this frequency range is dominated by the metamaterial structure and is largest in the vicinity where $\epsilon_1(\omega) \approx 1$.

This active metamaterial device was designed such that the LC resonant response occurs with the electric field polarized perpendicular to the connecting wires. This eliminates the Drude-like response that occurs when the THz electric field is parallel to the array of connecting wires¹³. Nonetheless, with the electric field applied parallel to the connecting wires, a response that changes with applied bias is still observed, as shown in Fig. 4a. The small transmitted intensity at low frequencies results from the Drude-like response of the connecting wires. Superimposed on this is a resonance at 1.25 THz from the metamaterial elements. This resonant response is associated with dipolar currents in the elements analogous to the higher-lying resonance when the electric field is perpendicular to the connecting wires. As such, the variation in $\epsilon_1(\omega)$ shown in Fig. 4b

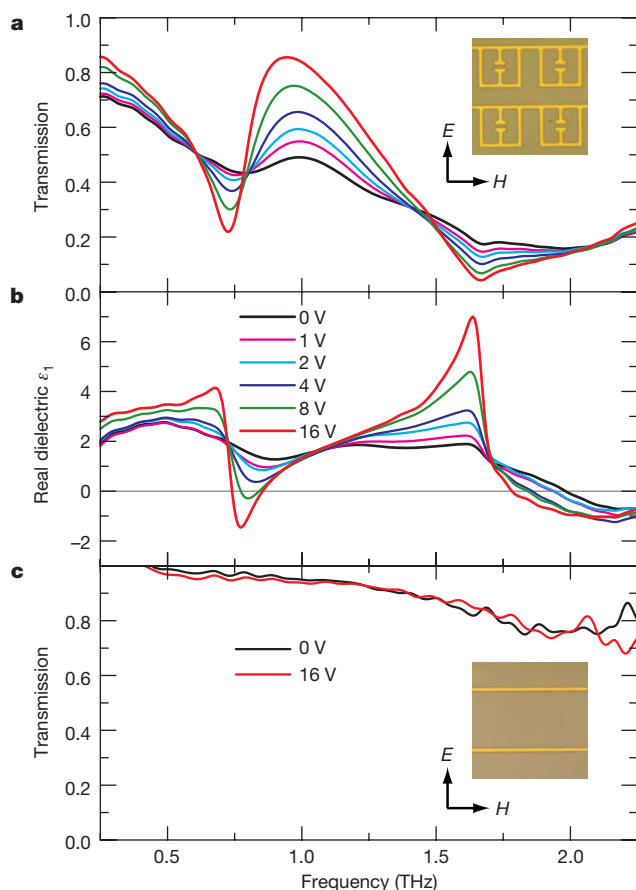


Figure 3 | Switching performance of the active THz metamaterial device as a function of gate voltage bias with the polarization of the THz electric field perpendicular to the connecting wires. a, Frequency-dependent transmitted intensity of THz radiation, and **b**, the corresponding permittivity for various reverse gate biases. **c**, THz transmission through a device with metamaterials removed, at reverse biases of 0 and 16 V. The insets show the polarization configuration of the THz electric and magnetic fields.

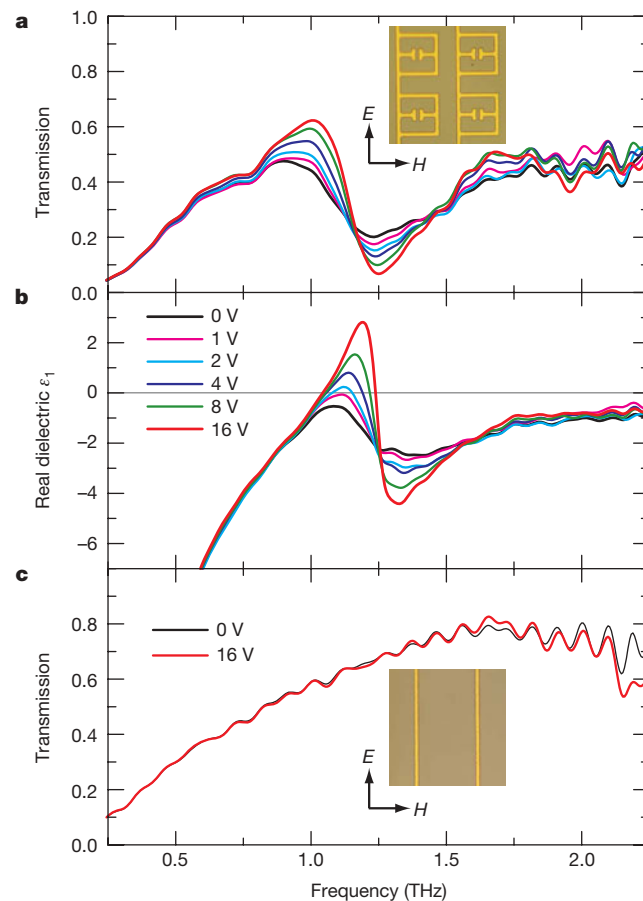


Figure 4 | Switching performance of the active THz metamaterial device as a function of gate voltage bias with the polarization of the THz electric field parallel to the connecting wires. a–c, As Fig. 3 but with the polarization of the THz electric field parallel to the connecting wires.

arises from substrate carriers screening the dipolar currents. An applied bias depletes the carriers, thus restoring the dipolar resonant response. Figure 4c shows the response for an array of parallel wires as the Schottky contact to the n-type substrate without the metamaterial elements. As expected, only a Drude-like response associated with the wires is observed. Thus without the resonant metamaterial elements and their critical dependence upon substrate properties, it is not possible to modulate the transmission with an applied bias. This further confirms the importance of the metamaterial elements in creating active THz devices.

Substrates typically used to fabricate planar metamaterial structures (for example, Si, GaAs, Teflon) are insulators and are essentially lossless at THz frequencies. In this case, the metamaterial structure and the substrate can be modelled as an equivalent LCR resonant circuit as shown in Fig. 1b without the variable resistor R_d . However, when the substrate is lossy (in our case this is a result of doping), the finite resistance at the split gap has to be considered. The equivalent circuit should be modified by attaching a variable resistor in parallel to the capacitor²⁷. In our device, the gate bias changes this resistance by depleting the free charge carriers and modifies the resonance strength. The fact that there are no significant shifts of the resonance frequencies indicates that the magnitude of the capacitance at the split gap is not strongly affected by the applied gate bias—rather, the capacitance is shunted.

Although the metamaterial structure presented here is a first generation device (that is, no optimization has been attempted), the performance as a THz modulator already exceeds current state-of-the-art electrical THz modulators—based on semiconductor structures¹⁰—by one order of magnitude on resonance, and operates at room temperature. Higher modulation efficiency for practical applications is expected to be achievable through device optimization, that is, by varying the doping concentration and/or the thickness of the doping layer. One problem with the current design is that high frequency modulation is not possible. We performed measurements of the THz intensity as a function of modulation frequency by applying a rectangular a.c. reverse gate bias alternating between 0 and 16 V. The large area ($5 \times 5 \text{ mm}^2$) of the metamaterial array results in a large overall device capacitance, yielding a maximum modulation frequency of several kilohertz (Supplementary Fig. 2). We anticipate significant increases in the modulation frequency from reducing the total capacitance and resistance of the device by, for example, patterning the n-GaAs regions of the substrate and/or using interdigitated contacts.

This demonstration of an active metamaterial device relied on electrically connecting the individual metamaterial elements. It is important to emphasize that this does not compromise the resonant LC response of the elements, and thereby provides additional design flexibility for metamaterials in general. The approach presented here for active THz metamaterials naturally extends to magnetically resonant metamaterials. Finally, consideration of the substrate or embedding environment also offers considerable flexibility in the design of active metamaterial devices at any frequency range.

Received 3 July; accepted 10 October 2006.

1. Veselago, V. G. The electrodynamics of substances with simultaneously negative values of ϵ and μ . *Sov. Phys. Usp.* **10**, 509–514 (1968).
2. Oliveira, F. *et al.* Neural network analysis of terahertz spectra of explosives and bio-agents. *Proc. SPIE* **5070**, 60–70 (2003).
3. Zimdars, D. Fiber-pigtailed terahertz time domain spectroscopy instrumentation for package inspection and security imaging. *Proc. SPIE* **5070**, 108–116 (2003).

4. Federici, J. F. *et al.* THz imaging and sensing for security applications—explosives, weapons and drugs. *Semicond. Sci. Technol.* **20**, S266–S280 (2005).
5. Barber, J. *et al.* Temperature-dependent far-infrared spectra of single crystals of high explosives using terahertz time-domain spectroscopy. *J. Phys. Chem. A* **109**, 3501–3505 (2005).
6. Liu, H.-B., Chen, Y., Bastiaans, G. J. & Zhang, X.-C. Detection and identification of explosive RDX by THz diffuse reflection spectroscopy. *Opt. Express* **14**, 415–423 (2006).
7. Kawase, K., Ogawa, Y., Watanabe, Y. & Inoue, H. Non-destructive terahertz imaging of illicit drugs using spectral fingerprints. *Opt. Express* **11**, 2549–2554 (2003).
8. Ferguson, B. & Zhang, X.-C. Materials for terahertz science and technology. *Nature Mater.* **1**, 26–33 (2002).
9. Köhler, R. *et al.* Terahertz semiconductor-heterostructure laser. *Nature* **417**, 156–159 (2002).
10. Kleine-Ostmann, T., Dawson, P., Pierz, K., Hein, G. & Koch, M. Room-temperature operation of an electrically driven terahertz modulator. *Appl. Phys. Lett.* **84**, 3555–3557 (2004).
11. Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. Composite medium with simultaneously negative permeability and permittivity. *Phys. Rev. Lett.* **84**, 4184–4187 (2000).
12. Shelby, R. A., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292**, 77–79 (2001).
13. Pendry, J. B., Holden, A. J., Stewart, W. J. & Youngs, I. Extremely low frequency plasmons in metallic mesostructures. *Phys. Rev. Lett.* **76**, 4773–4776 (1996).
14. Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microwave Theory Tech.* **47**, 2075–2084 (1999).
15. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
16. Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
17. Leonhardt, U. Optical conformal mapping. *Science* **312**, 1777–1780 (2006).
18. Zhang, S. *et al.* Experimental demonstration of near-infrared negative-index metamaterials. *Phys. Rev. Lett.* **95**, 137404 (2005).
19. Zhou, J. *et al.* Saturation of the magnetic response of split-ring resonators at optical frequencies. *Phys. Rev. Lett.* **95**, 223902 (2005).
20. Grigorenko, A. N. *et al.* Nanofabricated media with negative permeability at visible frequencies. *Nature* **438**, 335–338 (2005).
21. Yen, T. J. *et al.* Terahertz magnetic response from artificial materials. *Science* **303**, 1494–1496 (2004).
22. Padilla, W. J., Taylor, A. J., Highstrete, C., Lee, M. & Averitt, R. D. Dynamical electric and magnetic metamaterial response at terahertz frequencies. *Phys. Rev. Lett.* **96**, 107401 (2006).
23. Wang, K. & Mittleman, D. M. Metal wires for terahertz wave guiding. *Nature* **432**, 376–379 (2004).
24. Mendis, R. & Grischkowsky, D. THz interconnect with low-loss and low-group velocity dispersion. *IEEE Microwave Wireless Compon. Lett.* **11**, 444–446 (2001).
25. Kersting, R., Strasser, G. & Unterrainer, K. Terahertz phase modulator. *Electron. Lett.* **36**, 1156–1158 (2000).
26. Schurig, D., Mock, J. J. & Smith, D. R. Electric-field-coupled resonators for negative permittivity metamaterials. *Appl. Phys. Lett.* **88**, 041109 (2006).
27. Ikonen, P. & Tretyakov, S. Generalized permeability function and field energy density in artificial magnetics. Preprint at <http://arxiv.org/abs/physics/0602182> (2006).
28. Grischkowsky, D., Keiding, S., van Exter, M. & Fittinger, Ch Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. *J. Opt. Soc. Am. B* **7**, 2006–2015 (1990).
29. O'Hara, J. F., Zide, J. M. O., Gossard, A. C., Taylor, A. J. & Averitt, R. D. Enhanced terahertz detection via ErAs:GaAs nanoisland superlattices. *Appl. Phys. Lett.* **88**, 251119 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from the Los Alamos National Laboratory LDRD programme, and from the Center for Integrated Nanotechnologies.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.-T.C. (chenht@lanl.gov).

Gulf Stream density structure and transport during the past millennium

David C. Lund^{1†}, Jean Lynch-Stieglitz² & William B. Curry³

The Gulf Stream transports approximately 31 Sv (1 Sv = $10^6 \text{ m}^3 \text{ s}^{-1}$) of water^{1,2} and $1.3 \times 10^{15} \text{ W}$ of heat³ into the North Atlantic ocean. The possibility of abrupt changes in Gulf Stream heat transport is one of the key uncertainties in predictions of climate change for the coming centuries. Given the limited length of the instrumental record, our knowledge of Gulf Stream behaviour on long timescales must rely heavily on information from geologic archives. Here we use foraminifera from a suite of high-resolution sediment cores in the Florida Straits to show that the cross-current density gradient and vertical current shear of the Gulf Stream were systematically lower during the Little Ice Age (AD ~1200 to 1850). We also estimate that Little Ice Age volume transport was ten per cent weaker than today's. The timing of reduced flow is consistent with temperature minima in several palaeoclimate records⁴⁻⁹, implying that diminished oceanic heat transport may have contributed to Little Ice Age cooling in the North Atlantic. The interval of low flow also coincides with anomalously high Gulf Stream surface salinity¹⁰, suggesting a tight linkage between the Atlantic Ocean circulation and hydrologic cycle during the past millennium.

The Florida Current, the portion of the Gulf Stream that flows through the Straits of Florida, is characterized by a strong cross-current gradient in seawater density. This gradient reflects a flow that is in approximate geostrophic and hydrostatic balance, and it is proportional to the vertical shear in current velocity. Past changes in current shear can be estimated using two vertical profiles of seawater density, one from each side of the Florida Current. To reconstruct density profiles for the past millennium, we used foraminifera from sediment cores retrieved near Dry Tortugas and Great Bahama Bank (Fig. 1). The $\delta^{18}\text{O}$ of foraminifera ($\delta^{18}\text{O}_c$) depends on the temperature and the $\delta^{18}\text{O}$ of sea water ($\delta^{18}\text{O}_w$) in which they live. Because $\delta^{18}\text{O}_w$ is linearly related to salinity, and the difference between $\delta^{18}\text{O}_w$ and $\delta^{18}\text{O}_c$ is thermodynamically controlled, seawater density can be estimated using $\delta^{18}\text{O}_c$ (ref. 11; see Methods).

To estimate volume transport using the geostrophic technique, a reference velocity must be known or estimated to provide a constant of integration for the vertical shear profile. Typically, a level of no motion is assumed near the bottom of the flow. Today, near-zero velocities in this part of the Florida Straits occur below 800 m (refs 2, 12). We used a level of no motion at a water depth of 850 m, which yields a modern transport of 31 Sv (see Methods and Supplementary Information on transport calculation). This value is consistent with flow estimates obtained using current meter¹³⁻¹⁵, geostrophic¹⁶, cable voltage^{1,17}, dropsonde^{14,18}, and Pegasus acoustic profiler techniques².

The Florida Current density time series are based on over 3,000 stable isotopic measurements of foraminifera calcifying at water depths ranging from the surface to 750 m and ranging in age from ~0 to 1,100 yr ago (Fig. 2). Planktonic foraminiferal $\delta^{18}\text{O}_c$ was used

to constrain the density of the surface mixed layer. The annual average mixed layer in this region is approximately 50 m (ref. 19), similar to the depth habitat of *Globigerinoides ruber*^{20,21}. Near Dry Tortugas, planktonic $\delta^{18}\text{O}_c$ was constant (within 95% confidence limits) from 2,000 to 900 yr before present (BP), but then increased by 0.3‰ ($\Delta\sigma_t = \sim 0.5$) from 900 to 400 yr BP, with the largest shift after 600 yr BP (Fig. 2a). On the Bahamas side of the Florida Straits, $\delta^{18}\text{O}_c$ was stable from 2,000 to 250 yr BP, and then increased 0.2‰ ($\Delta\sigma_t = \sim 0.3$) from 250 yr BP to present (Fig. 2b). Surface water $\delta^{18}\text{O}_c$ variability at each site was driven primarily by changes in $\delta^{18}\text{O}_w$ (ref. 10).

The benthic $\delta^{18}\text{O}_c$ records show that changes in Florida Current density structure occurred during the past millennium (Fig. 2c, d). From ~1,100 to 900 yr BP, $\delta^{18}\text{O}_c$ increased by 0.1‰ ($\Delta\sigma_t = \sim 0.1$) along the Great Bahama Bank, yet there was little change at the Dry Tortugas sites, yielding a decrease in cross-current density gradient and vertical current shear. A further reduction in shear occurred from 900 to 200 yr BP, because $\delta^{18}\text{O}_c$ decreased by 0.1‰ at intermediate water depths near Dry Tortugas, while remaining nearly constant along the Great Bahama Bank. From 200 yr BP to the present, benthic $\delta^{18}\text{O}_c$ at all the Bahamas sites decreased by 0.1‰ ($\Delta\sigma_t = \sim 0.1$), with no comparable change at Dry Tortugas. This density gradient increase reflects enhanced vertical shear in the Florida Current during the past 200 yr.

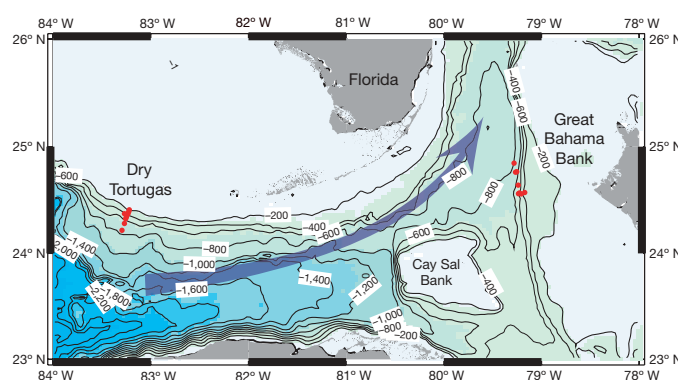


Figure 1 | Bathymetric map of Florida Straits showing core locations (red circles). Near Dry Tortugas, we used seven cores at five separate water depths, spanning 200 to 750 m (Supplementary Table 1). On the Great Bahama Bank, a total of seven cores were used at four water depths from 260 to 700 m (Supplementary Table 2). Contours indicate water depth in metres. All cores reported here, except W167-79GGC (ref. 30), were collected in January 2002 aboard the R/V *Knorr*. Approximately 28 Sv of transport occurs between Cay Sal Bank and Florida (blue arrow)^{2,16,17}, while ~2 Sv flows through the Santaren channel between Cay Sal Bank and Great Bahama Bank².

¹Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program in Oceanography, Woods Hole, Massachusetts 02543, USA. ²School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. ³Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [†]Present address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

Using the foraminifera-based seawater density data from Fig. 2, we estimate that Florida Current volume transport varied by 3 ± 1 Sv during the past millennium, or about 10% of the total flow (Fig. 3b). Compared to the 10–15 Sv estimated flow reduction for the Last Glacial Maximum²², the Florida Current is characterized by modest variations in volume transport during the latest Holocene. However, there is a distinct interval of low flow from ~700 to 100 yr BP—the approximate time of maximum cooling in several North Atlantic temperature records^{4–9}. The similar timing of changes in North Atlantic temperature and Florida Current strength suggests that reduced northward oceanic heat transport may have contributed to cooler temperatures during the Little Ice Age. Heat flux is a function of both volume transport and the temperature difference between the Florida Current and water flowing southward out of the North Atlantic³, so temperature reconstructions from these flows will be necessary to confirm past changes in oceanic heat transport.

In addition to total volume transport, the $\delta^{18}\text{O}_c$ -based density records yield information on the variability in transport with depth. The Florida Current consists of two primary components, one from the South Atlantic that is fresh (salinity < 36) and the other from the North Atlantic subtropical gyre that is salty (salinity > 36). The fresh component is found at the surface (>24 °C) and bottom (7–12 °C) of the Florida Current, accounts for ~14 Sv of the total transport, and acts as the northward-flowing upper limb of the meridional overturning circulation (MOC)²³. The deepest portion of the

Florida Current contains Antarctic Intermediate Water that has been modified through mixing with saltier water of the subtropical gyre. The wind-driven subtropical gyre component sits between the fresh South Atlantic water at the surface and deepest parts of the Florida Current, is characterized by intermediate temperatures (12–24 °C), and accounts for ~17 Sv of the total flow²⁴.

Reduced transport during the Little Ice Age is a function of lower flow throughout the water column, but the largest negative anomalies occur at depths from 100 to 500 m (or $24 < \sigma_t < 27$) (Fig. 3a). This is the density range associated with waters from the North Atlantic subtropical gyre, so it appears that the wind-driven component of the Florida Current was primarily responsible for low total transport during the Little Ice Age. The intervals 0–100 yr BP and 1,000–1,100 yr BP are characterized by higher transport, but they differ in that the modern anomaly spans the entire water column, whereas at 1,100 yr BP it is concentrated above 500 m. This pattern suggests that the flow of upper Antarctic Intermediate Water was weaker at 1,100 yr BP than it is today.

Flow anomalies below 500 m water depth are small for two reasons. First, we assume a level of no motion at 850 m and therefore departures from this reference point are minimized. Second, when contoured in terms of Sverdrups, the anomalies below 500 m appear small because total volume transport in this depth range is low relative to transport above 500 m. If we instead contour the percentage change in flow relative to the long-term mean, we observe a transport

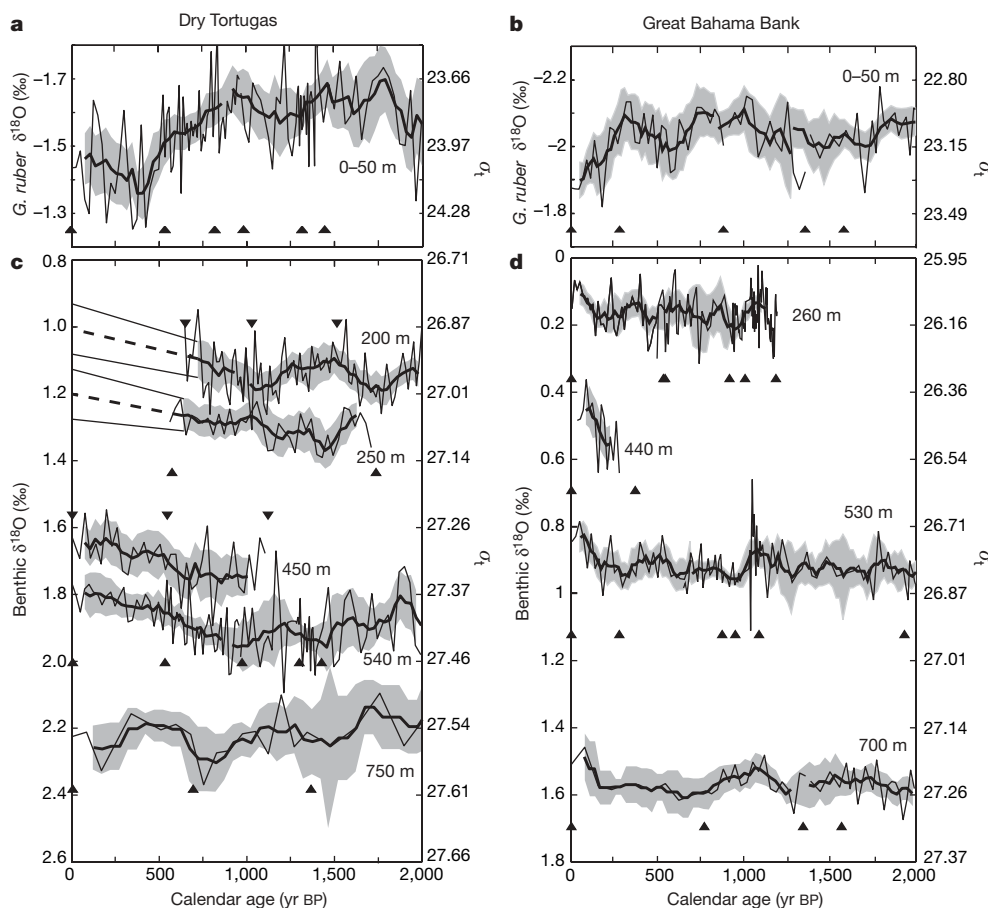


Figure 2 | Foraminiferal $\delta^{18}\text{O}_c$ data for Florida Straits cores. Plotted are the average $\delta^{18}\text{O}_c$ value at each stratigraphic level (thin black line), the running mean (thick black line), and the 95% confidence limit on the running mean (shaded area) (see Methods). Triangles represent calendar-corrected radiocarbon ages (Supplementary Tables 1 and 2). **a**, *G. ruber* $\delta^{18}\text{O}_c$ for Dry Tortugas. **b**, *G. ruber* $\delta^{18}\text{O}_c$ for Great Bahama Bank. **c**, Benthic $\delta^{18}\text{O}_c$ for Dry Tortugas. To fill the gap between 500 yr BP and today (at 200 and 250 m), we interpolated between σ_t values based on CTDs and core tops (Supplementary Fig. 4). The resulting trend (dashed line) is similar to that in

deeper cores, suggesting that it is a reasonable approximation of density for this time interval. The maximum $\delta^{18}\text{O}_c$ uncertainty for these sites is $\pm 0.1\text{‰}$ ($\pm 3\sigma$) ($\delta^{18}\text{O}_c$ must always be larger at 250 m than at 200 m), equivalent to a 95% confidence limit of $\pm 0.07\text{‰}$. Uncertainty between today and 500 yr BP was estimated by interpolation between the $\pm 0.07\text{‰}$ envelope and core top uncertainties. **d**, Benthic $\delta^{18}\text{O}_c$ for Great Bahama Bank. The time series at 440 m was not used in the transport calculation because it is too short, but the trend of decreasing $\delta^{18}\text{O}_c$ over the past 200 yr is consistent with the other benthic records from the Bahamas.

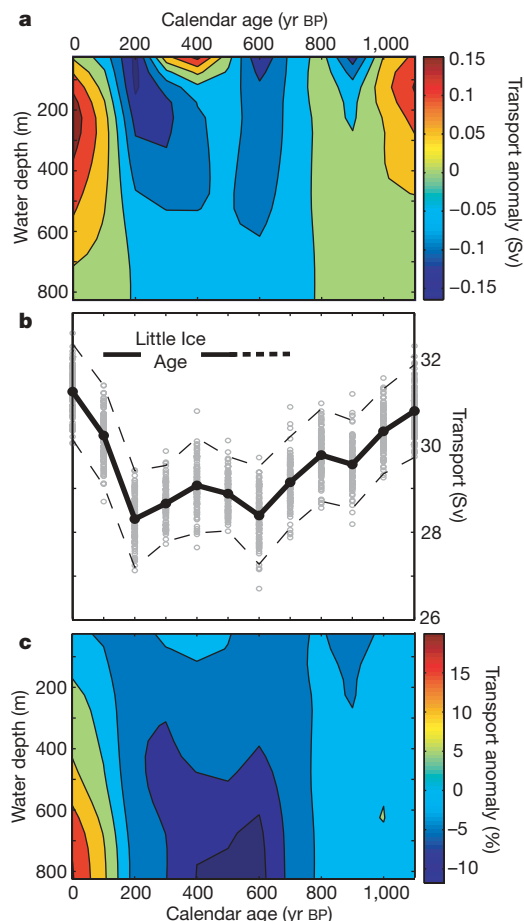


Figure 3 | Transport reconstruction for the Florida Current. **a**, Flow anomaly versus depth, defined as transport at a given time minus average transport over the period 0–1100 yr BP. Negative transport anomalies occur during the Little Ice Age and are largest in the depth range occupied by water from the North Atlantic subtropical gyre. **b**, Estimated total transport 0–1100 yr BP. The most probable transport values (thick black line) are based on mean densities in Fig. 2. The grey circles represent transport estimates based on density values randomly sampled from within the error envelopes in Fig. 2, while the dashed lines represent the 95% confidence limit for the transport calculation (see Methods). **c**, Flow anomaly versus depth, defined as the percentage change in transport relative to the mean value at a given depth from 0–1000 yr BP. The largest changes occur near the bottom of the Florida Current, the depth range associated with Antarctic Intermediate Water.

reduction of 15–25% during the Little Ice Age relative to today below 500 m water depth (Fig. 3c). This suggests the MOC also played a role in the Little Ice Age transport anomaly. Using only data from the Straits of Florida, it is not possible to quantify the relative contributions of the wind-driven gyre and MOC to the flow variability of the past millennium. Estimating changes in these components will require transport reconstructions from additional sites in the Atlantic.

Although the Little Ice Age is best-known as a time of cooler temperatures and alpine glacier advances in the Northern Hemisphere, it was also characterized by anomalously dry conditions in Central and South America^{25,26} and high surface salinity in the Florida Current (Fig. 4). A comparison of Florida Current surface $\delta^{18}\text{O}_w$ to the Cariaco Basin aridity record implies that there is a strong climatic link between the two¹⁰. As originally suggested by ref. 25, the drier conditions probably reflect a southward shift of the Atlantic Inter-Tropical Convergence Zone (ITCZ). On interannual to decadal timescales, small ($<1^\circ\text{C}$) shifts in the cross-equatorial Atlantic sea surface temperature (SST) gradient can cause the mean

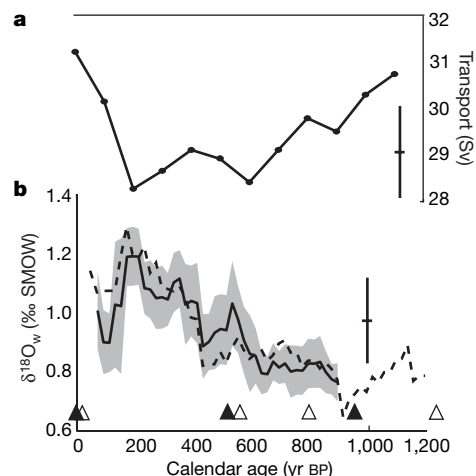


Figure 4 | Comparison of Florida Current transport and surface water $\delta^{18}\text{O}_w$ reconstructions. **a**, Transport 0–1100 yr BP. The vertical bar represents the average 95% error envelope from Fig. 3 (± 1 Sv). **b**, Surface water $\delta^{18}\text{O}_w$ estimates based on $\delta^{18}\text{O}_c$ and Mg/Ca analyses of planktonic foraminifera (*G. ruber*) in two Dry Tortugas cores¹⁰. Lines represent 100-yr running mean values for $\delta^{18}\text{O}_w$ (solid line, KNR166-2-62MC; dashed line, W167-79GGC). The shaded area represents the 95% confidence limit for 62MC. For clarity, the average 95% confidence limit for 79GGC ($\pm 0.15\text{‰}$) is given as a vertical bar. The $\delta^{18}\text{O}_w$ error estimates are based on the combined uncertainty of Mg/Ca and $\delta^{18}\text{O}_c$ analyses¹⁰. Symbols indicate calendar-calibrated radiocarbon control points (solid triangles, KNR166-2-62MC; open triangles, W167-79GGC).

ITCZ position to shift more than 1,000 km (ref. 27). We speculate that reduced northward oceanic heat transport during the Little Ice Age altered the tropical Atlantic SST field and triggered southward movement of the ITCZ. This scenario is consistent with evidence for reduced volume transport and high surface salinities in the Straits of Florida (Fig. 4).

Control simulations of the HadCM3 coupled ocean–atmosphere model yield a centennial-scale climate oscillation involving the MOC and the ITCZ²⁸. A reduction in the MOC of 2 Sv produces a $\sim 0.1^\circ\text{C}$ cross-equatorial Atlantic SST contrast that forces southward ITCZ migration and yields positive salinity anomalies in the tropical North Atlantic. The simulated anomalies advect to the high-latitude North Atlantic via the Gulf Stream, and act to enhance the MOC by increasing surface ocean density. The resulting increase in northward oceanic heat transport alters the cross-equatorial Atlantic SST gradient, causing the ITCZ to migrate northwards, producing a negative salinity anomaly in the tropical North Atlantic. As this low-salinity water advects to deep-water convection sites in the Greenland and Nordic Seas, the strength of the MOC diminishes, thus completing the oscillation.

We speculate that a similar coupling occurred during the past millennium, where changes in Gulf Stream salinity acted as a negative feedback on volume transport by modulating the strength of the MOC through buoyancy forcing in the high-latitude Atlantic. It is also possible that southward ITCZ migration, perhaps triggered by low solar irradiance¹⁰, reduced windstress curl over the North Atlantic subtropical gyre and therefore northward oceanic heat transport. In either case, it appears that the Gulf Stream played a key role in the Little Ice Age, perhaps as a feedback in an oscillation internal to the climate system or as an amplifier of externally forced variability.

METHODS

Age control. Age control is provided by multiple radiocarbon dates on planktonic foraminifera in each core (see Supplementary Tables 1 and 2). All raw radiocarbon ages were converted to calendar ages assuming a surface ocean reservoir age of 400 yr (see Supplementary Information on age control). Age models were created by linear interpolation between ^{14}C control points. Average

sedimentation rates for the Dry Tortugas cores range from 11 to 66 cm kyr⁻¹. For the Great Bahama Bank, average sedimentation rates are higher, ranging from 20 to 350 cm kyr⁻¹.

$\delta^{18}\text{O}_\text{c}$ analyses. Stable isotope analyses for the Florida Straits cores were based on *Cibicides* and *Planulina* species. These species follow the fractionation for inorganic calcite precipitation determined in ref. 29, such that the $\delta^{18}\text{O}_\text{calcite} - \delta^{18}\text{O}_\text{seawater}$ difference follows a thermodynamic slope of $\sim 0.2\text{‰ } ^\circ\text{C}^{-1}$ (see Supplementary Information on sampling strategy). Stable-isotope analyses were performed on a Finnigan MAT 253 coupled to a Kiel III carbonate device. Calibration to the Vienna Pee-Dee Belemnite (VPDB) scale was made using NBS-19 ($\delta^{18}\text{O} = -2.20\text{‰}$ and $\delta^{13}\text{C} = 1.95\text{‰}$). Long-term reproducibility (1σ) of NBS-19 ($n = 461$) for this mass spectrometer system is $\pm 0.08\text{‰}$ for $\delta^{18}\text{O}$ and $\pm 0.04\text{‰}$ for $\delta^{13}\text{C}$.

The 95% confidence limits for the running-mean $\delta^{18}\text{O}_\text{c}$ values in Fig. 2 are based on the Student's t -distribution (upper limit = $M + t(\sigma/\sqrt{n})$; lower limit = $M - t(\sigma/\sqrt{n})$; where M is the running mean, t is the two-tailed t -value, σ is the standard deviation of $\delta^{18}\text{O}_\text{c}$ analyses for a given time window, and n is the number of analyses). The length of the smoothing window for each core was chosen such that each running-mean $\delta^{18}\text{O}_\text{c}$ value is based on an average 15–25 individual $\delta^{18}\text{O}_\text{c}$ analyses (for $\alpha = 0.05$ and $n = 20$, $t = 2.09$). All of the Dry Tortugas cores were smoothed using a 150-yr window, except the lowest-sedimentation-rate core (at 750 m water depth), where we used a 250-yr window. The Great Bahama Bank $\delta^{18}\text{O}_\text{c}$ time series were smoothed with a 100-yr window, except at 700 m water depth, where a 150-yr window was used.

Density calibration. We quantified the relationship between foraminiferal $\delta^{18}\text{O}$ ($\delta^{18}\text{O}_\text{c}$) and seawater σ_t by: (1) estimating seawater $\delta^{18}\text{O}$ ($\delta^{18}\text{O}_\text{w}$) for conductivity–temperature–depth (CTD) data taken at the core sites using the $\delta^{18}\text{O}_\text{w}$ – σ_t relationship from North Atlantic thermocline waters; (2) calculating the predicted $\delta^{18}\text{O}_\text{c}$ of foraminifera precipitated in equilibrium with this water; (3) determining σ_t values for the CTD data, and finally (4) creating a regression between σ_t and $\delta^{18}\text{O}_\text{c}$ (see Supplementary Information on density calibration). The resulting polynomial fit was used to estimate past changes in seawater density. Density values based on foraminifera from the most recent (core top) sediment in each core are very similar to density values determined using temperature and salinity data from CTD casts (see Supplementary Fig. 4).

Transport calculation. Transport estimates were made using the thermohaline anomaly, which is the primary component of specific volume anomaly in the upper 1,000 m of the ocean and can be calculated directly from σ_t (see Supplementary Information on transport calculation). We used a level of no motion (850 m) that yields a core top transport of 31 Sv, similar to the modern instrumental value. This level of no motion is consistent with current meter data that show near-zero velocities between Cay Sal Bank and Florida at 800–1,000 m (ref. 2). Offshore of Key West, the level of no motion is deeper ($\sim 1,000$ m) but all velocities below 600 m are less than 5 cm s^{-1} (ref. 12). In the much shallower Santaren Channel, near-zero flow velocities occur near 600 m (ref. 2). A level of no motion at 900 m yields a transport of 33 Sv, which is higher than published values of annual average flow for the Florida Current at this location.

We estimated uncertainty in the transport calculation using a Monte Carlo approach. For each time interval, transport was calculated using σ_t values randomly sampled from the observed probability distribution in each core (Fig. 2), which reflects the analytical and sampling uncertainty of the $\delta^{18}\text{O}_\text{c}$ analyses. We repeated this process 100 times at each 100-yr increment from 0 to 1,100 yr BP (Fig. 3; grey circles). The 95% error envelope (Fig. 3; dashed lines), defined here as $\pm 2\sigma$ of the Monte Carlo estimates, averages ± 1 Sv. Reasonable changes in the $\delta^{18}\text{O}_\text{w}$ – σ_t relationship have a small (<0.5 Sv) impact on transport (see Supplementary Information on $\delta^{18}\text{O}_\text{w}$ – σ_t transport error).

Received 14 March; accepted 19 September 2006.

1. Baringer, M. O. & Larsen, J. C. Sixteen years of Florida Current transport at 27°N. *Geophys. Res. Lett.* **28**, 3179–3182 (2001).
2. Leaman, K. D. *et al.* Transport, potential vorticity, and current/temperature structure across Northwest Providence and Santaren Channels and the Florida Current off Cay Sal Bank. *J. Geophys. Res.* **100** (C5), 8561–8569 (1995).
3. Larsen, J. C. Transport and heat flux of the Florida Current at 27°N derived from cross-stream voltages and profiling data: theory and observations. *Phil. Trans. R. Soc. Lond. A* **338**, 169–236 (1992).
4. Lamb, H. *Climate, History and the Modern World* edn 2 (Routledge, London/New York, 1995).
5. Keigwin, L. D. The Little Ice age and Medieval Warm Period in the Sargasso Sea. *Science* **274**, 1504–1508 (1996).
6. Dahl-Jensen, D. *et al.* Past temperatures directly from the Greenland ice sheet. *Science* **282**, 268–271 (1998).

7. deMenocal, P., Ortiz, J., Guilderson, T. & Sarnthein, M. Coherent high- and low-latitude climate variability during the Holocene warm period. *Science* **288**, 2198–2202 (2000).
8. Marchitto, T. M. & deMenocal, P. B. Late Holocene variability of upper North Atlantic Deep Water temperature and salinity. *Geochim. Geophys. Geosyst.* **4**, 1100, doi:10.1029/2003GC000598 (2003).
9. Esper, J., Cook, E. R. & Shweingruber, F. H. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* **295**, 2250–2253 (2002).
10. Lund, D. C. & Curry, W. B. Florida Current surface temperature and salinity variability during the last millennium. *Paleoceanography* **21**, PA2009, doi:10.1029/2005PA001218 (2006).
11. Lynch-Stieglitz, J., Curry, W. B. & Slowey, N. A geostrophic estimate for the Florida Current from the oxygen isotope composition of benthic foraminifera. *Paleoceanography* **14**, 360–373 (1999).
12. Hamilton, P., Larsen, J. C., Leaman, K. D., Lee, T. N. & Waddell, E. Transports through the Straits of Florida. *J. Phys. Oceanogr.* **35**, 308–322 (2005).
13. Pillsbury, J. E. The Gulf Stream—A description of the methods employed in the investigation, and the results of the research. In *Report of the Superintendent of the U. S. Coast and Geodetic Survey* 173–184 (U. S. Coast and Geodetic Survey, Washington DC, 1890).
14. Schmitz, W. J. & Richardson, W. S. On the transport of the Florida current. *Deep-Sea Res.* **15**, 679–693 (1968).
15. Schott, F. A., Lee, T. N. & Zantopp, R. Variability and structure of the Florida Current in the period range of days to seasonal. *J. Phys. Oceanogr.* **18**, 1209–1230 (1988).
16. Montgomery, R. B. Transport of the Florida Current off Habana. *J. Mar. Res.* **IV**, 198–220 (1941).
17. Wertheim, G. K. Studies of the electric potential between Key West, Florida, and Havana, Cuba. *Trans. AGU* **35**, 872–882 (1954).
18. Niiler, P. P. & Richardson, W. S. Seasonal variability of the Florida Current. *J. Mar. Res.* **31**, 144–167 (1973).
19. Carton, J. A. & Giese, B. S. SODA version 1.4.2: A reanalysis of ocean climate. *J. Geophys. Res.* (submitted); preprint at (<http://www.atmos.umd.edu/~carton/pdfs/carton&giese05.pdf>) (2005).
20. Deuser, W. G. Seasonal variations in isotopic composition and deep-water fluxes of the tests of perennially abundant planktonic foraminifera of the Sargasso Sea: results from sediment-trap collections and their paleoceanographic significance. *J. Foraminiferal Res.* **17**, 14–27 (1987).
21. Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**, 1050, doi:10.1029/2002PA000846 (2003).
22. Lynch-Stieglitz, J., Curry, W. B. & Slowey, N. Weaker Gulf Stream in the Florida Straits during the Last Glacial Maximum. *Nature* **402**, 644–648 (1999).
23. Schmitz, W. J. & Richardson, P. On the sources of the Florida Current. *Deep-Sea Res.* **38** (suppl.), 379–409 (1991).
24. Schmitz, W. J., Luyten, J. R. & Schmitt, R. W. On the Florida Current T/S envelope. *Bull. Mar. Sci.* **53**, 1048–1065 (1993).
25. Haug, G., Hughen, K. A., Sigman, D. M., Peterson, L. C. & Rohl, U. Southward migration of the Inter-Tropical Convergence Zone through the Holocene. *Science* **293**, 1304–1308 (2001).
26. Hodel, D. A. *et al.* Climate change on the Yucatan Peninsula during the Little Ice Age. *Quat. Res.* **63**, 109–121 (2005).
27. Chiang, J., Kushnir, Y. & Giannini, A. Deconstructing Atlantic Intertropical Convergence Zone variability: Influence of the local cross-equatorial sea surface temperature gradient and remote forcing from the eastern equatorial Pacific. *J. Geophys. Res.* **107**, doi:10.1029/2000JD000307 (2002).
28. Vellinga, M. & Wu, P. Low-latitude freshwater influence on centennial variability of the Atlantic thermohaline circulation. *J. Clim.* **17**, 4498–4511 (2004).
29. Kim, S. & O'Neil, J. Equilibrium and non-equilibrium oxygen isotope effects in synthetic carbonates. *Geochim. Cosmochim. Acta* **61**, 3461–3475 (1997).
30. Lund, D. C. & Curry, W. B. Late Holocene variability in Florida Current surface density: Patterns and possible causes. *Paleoceanography* **19**, PA4001, doi:10.1029/2004PA001008 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank O. Marchal, L. Keigwin, D. Oppo and J. McManus for suggestions. We also thank D. Ostermann, M. Jeglinski, P. Cerulli, S. Thorrold and S. Birdwhistle for technical support. We are grateful to the WHOI core lab for sample collection and archiving, the captain and crew of the R/V *Knorr*, and to the Sea Education Association for access to their vessel *Westward*. This work was supported by the US National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.C.L. (dlund@gps.caltech.edu).

Petrology and thermal structure of the Hawaiian plume from Mauna Kea volcano

Claude Herzberg¹

There is uncertainty about whether the abundant tholeiitic lavas on Hawaii are the product of melt from peridotite or pyroxenite/eclogite rocks^{1,2}. Using a parameterization of melting experiments on peridotite³ with glass analyses from the Hawaii Scientific Deep Project 2 on Mauna Kea volcano¹, I show here that a small population of the core samples had fractionated from a peridotite-source primary magma. Most lavas, however, differentiated from magmas that were too deficient in CaO and enriched in NiO (ref. 2) to have formed from a peridotite source. For these, experiments indicate that they were produced by the melting of garnet pyroxenite, a lithology that had formed in a second stage by reaction of peridotite with partial melts of subducted oceanic crust². Samples in the Hawaiian core are therefore consistent with previous suggestions that pyroxenite occurs in a host peridotite, and both contribute to melt production^{2,4}. Primary magma compositions vary down the drill core, and these reveal evidence for temperature variations within the underlying mantle plume. Mauna Kea magmatism is represented in other Hawaiian volcanoes, and provides a key for a general understanding of melt production in lithologically heterogeneous mantle.

Variations in MgO and SiO₂ for glasses¹ and whole rocks⁵ from Mauna Kea obtained from the Hawaii Scientific Drilling Project 2 (HSDP2) are summarized in Fig. 1a. All lavas are tholeiitic, but there are three main populations of data: a low- and a high-silica population, both of which are uniformly low in CaO, and a small population with low SiO₂ and high CaO and K₂O. A fourth group at 2,233 metres below sea level (m.b.s.l.) in the drill core is similar to the low-SiO₂ type except that it is higher in alkalis¹. Within the HSDP2 core, the low- and high-SiO₂ lavas are the most abundant, and they are observed in both whole rocks and glasses. The high-CaO types are more rarely found as glasses at 1800 mbsl in the drill core, and a few surface whole rock samples^{6,7}.

The less-common lavas with the highest CaO contents have the properties of forming from a mantle peridotite source. A primary magma is a partial melt of a mantle source, and several compositions are shown in Fig. 1b and given in Table 1. These have been calculated using a petrological method discussed in Herzberg and O'Hara⁸ and summarized in the footnote to Table 1. Peridotite-source primary magmas for Mauna Kea are very similar to those for other picrites and komatiites. All have about 10% CaO, 17–19% MgO, and 46–47% SiO₂, but Mauna Kea differs in having much larger abundances of TiO₂, K₂O, and all other incompatible elements. The 10% CaO is consistent with the new parameterization of CaO shown in Fig. 1b (Supplementary Figs 1–3). That is, the CaO contents of accumulated fractional melts of mantle peridotite do not change much over a wide range of initial and final melting pressures within the garnet lherzolite stability field. Although this result was modelled from experiments on a fertile peridotite³, there is some indication that it applies also to more depleted compositions. Gorgona komatiites melted from a

source with long-term light rare-earth elements depletion⁹ (that is, $\varepsilon_{\text{Nd}} = +6$ –10), indicating that the peridotite source might also be depleted in CaO and Al₂O₃; however, primary magmas consistently display CaO contents of about 10% (Table 1).

The CaO contents of the commoner low- and high-SiO₂ parental magmas were estimated to be about 8.6 and 8.9% CaO, respectively¹. Parental magmas are the most magnesian liquids that can be inferred from erupted lavas, and they are modified to some extent by fractional crystallization and/or mixing and crystal accumulation. Their compositions have been inferred by calculating liquid compositions that would crystallize Fo_{90.5} olivine, the maximum observed¹. These parental magmas have significantly less than 10% CaO in peridotite partial melts (Fig. 1b). Therefore, a normal peridotite source is inconsistent with these low-CaO contents (see Supplementary Information), a conclusion that can also be extended to most shield-building tholeiitic lavas from Hawaii. This result is not compromised by augite crystallization, which lowers CaO only when the parental magma evolves by olivine fractionation to about 7.5% MgO. A pyroxenite source was an alternative favoured by Sobolev *et al.*² because the Ni contents of most Hawaiian parental magmas and their crystallizing olivines are higher than expected for a peridotite source.

A pyroxenite model for the production of the voluminous low- and high-SiO₂ lavas is shown in Fig. 2. These are projections of lava compositions and phase diagrams that have been constructed from melting experiments (Supplementary Information). Liquids (L) of interest form an array of compositions (similar to Hawaiian primary magmas) that project along a line called a cotectic because they co-crystallize orthopyroxene, clinopyroxene and garnet [L+opx+cpx+gt]. The pyroxene–garnet plane is a thermal divide that separates cotectic liquids on each side from being derived from each other either by partial melting or crystallization^{10–12}. To visualize the thermal divide, these liquid compositions are projected to and from diopside into the plane olivine–quartz–CATS (calcium Tschermak's pyroxene), a Kogiso diagram¹² (Fig. 2a). Pyroxene and garnet compositions project along the line opx–CATS. The cotectic [L+opx+cpx+gt] consists of two segments separated by the thermal divide. The first segment is defined by liquids that become increasingly olivine-rich as temperature decreases, in blue in Fig. 2; the compositions of these liquids are well-known at 3 GPa, but only moderately constrained at 4 GPa (Supplementary Information). The second segment is defined by liquids that become increasingly SiO₂-rich as temperature decreases, in red in Fig. 2; the projected locations in Fig. 2 have been interpolated from experiments on natural MORB compositions crystallizing [L+cpx+gt+quartz or coesite]^{13,14} (Supplementary Information). There is some confidence about this interpolation from theoretical considerations (Supplementary Information), but calibration by additional experimental work is needed.

¹Department of Geological Sciences Rutgers University Piscataway, New Jersey 08854, USA.

Table 1 | Estimated compositions of peridotite-source primary magmas for Mauna Kea and other occurrences

	Mauna Kea samples				Gorgona	Baffin Island	West Greenland
	SR0684-8.95 (ref. 1)	SR0684-9.00 (ref. 1)	KI-8 (ref. 6)	C-209 (ref. 7)			
SiO ₂	46.3	46.4	46.47	45.84	46.1	46.5	46.9
TiO ₂	2.01	2.11	1.78	1.83	0.56	0.99	1.29
Al ₂ O ₃	9.93	9.93	8.6	9.9	11.7	12.1	11.6
Cr ₂ O ₃	-	-	0.26	-	0.16	0.09	0.05
Fe ₂ O ₃	0.93	0.90	1.24	1.23	1.18	0.99	1.82
FeO	10.4	10.1	10.3	10.32	10.1	9.7	9.4
MnO	0.18	-	0.16	0.19	0.18	0.17	0.21
MgO	17.8	18.1	19.0	18.27	18.8	17.9	17.1
CaO	10.0	10.0	10.1	10.12	10.0	10.1	9.9
Na ₂ O	1.72	1.69	1.73	1.53	1.04	1.30	1.56
K ₂ O	0.56	0.52	0.12	0.42	0.03	0.04	0.09
NiO	-	-	0.079	0.084	0.112	0.110	0.081
P ₂ O ₅	0.234	0.202	0.186	0.284	-	-	-
<i>k_d</i>	0.298	0.298	0.299	0.296	0.300	0.301	0.303
Olivine Mg#	91.1	91.4	91.6	91.4	91.7	91.6	91.5
<i>T</i> (°C)	1,405	1,411	1,428	1,414	1,425	1,407	1,391

Primary magmas are estimated with a hybrid forward and inverse model discussed in ref. 8. The forward model component identifies all potential primary magma compositions as functions of melt fraction for an assumed peridotite source and melting mechanism, based on a parameterization of experimental data. These are displayed in projections and FeO–MgO plots, and accumulated fractional melting is the mechanism assumed here. The inverse component selects a derivative liquid in an erupted lava suite into which olivine is incrementally added; this provides an array of possible primary magma composition. The potential primary magmas in both forward and inverse models are compared, and common melt fractions in both projection and FeO–MgO plots are used to seek a unique primary magma solution. Primary magma solutions are usually the same for both fertile and depleted peridotite sources with FeO = 8 wt%. Olivine is added to the four Mauna Kea samples shown. The Fe–Mg exchange coefficient *k_d* for olivine–L and Mg# for olivine are calculated using the method of Toplis²⁹ and FeO/ΣFe = 0.90. *T* (°C) is the olivine liquidus temperature calculated at 1 atmosphere using the method of Beattie³⁰. Primary magmas for komatiites from Gorgona komatiites, and picrites from Baffin Island and West Greenland are given for comparison⁹.

HSDP2 glasses with >7.5% MgO exhibit no signs of augite or plagioclase fractionation. To these samples olivine was added until the liquid reached the cotectic [L+opx+cpx+gt], yielding primary pyroxenite-source magmas. Many of these are very similar to parental magmas that have been independently estimated with Fo_{90.5} olivine as a constraint¹, which suggests no change in primary magma

composition during transit from the mantle to the crust. Representative compositions are given in Table 2. Figure 2 shows that high-SiO₂ parental and primary magmas project to the SiO₂-rich side of the thermal divide. In contrast, low-SiO₂ parental and primary magmas project to the olivine-rich side of the divide. There is a clear separation of these two populations of data, indicating that they are partial melts of garnet pyroxenite formed on the cotectic [L+opx+cpx+gt]. Whole-rock data are similar, but they are less separated than the glasses. This work supports models of pyroxenite within a peridotite matrix for Hawaii, and melt production from both sources^{2,4} (Figs 1b, 2).

The thermal divide predicts primary magma compositions on each side will become increasingly similar with increasing temperature and, indeed, this is seen in Fig. 3a for Na₂O, P₂O₅ and SiO₂ abundances from the glass data. This agreement between theory and observation is a positive test of the importance of pyroxenite melting along the cotectic [L+opx+cpx+gt]. Some of the scatter in the trends might be a reflection of original pyroxenite source heterogeneity. In contrast, whole rocks exhibit no such behaviour (Fig. 3b) for Na₂O or any other incompatible element, although both SiO₂ populations generally project on each side of the thermal divide. Some of the whole-rock variability arises by addition and subtraction of augite and plagioclase³, which has been filtered out (Supplementary Information) to obtain as much of the partial melting signature

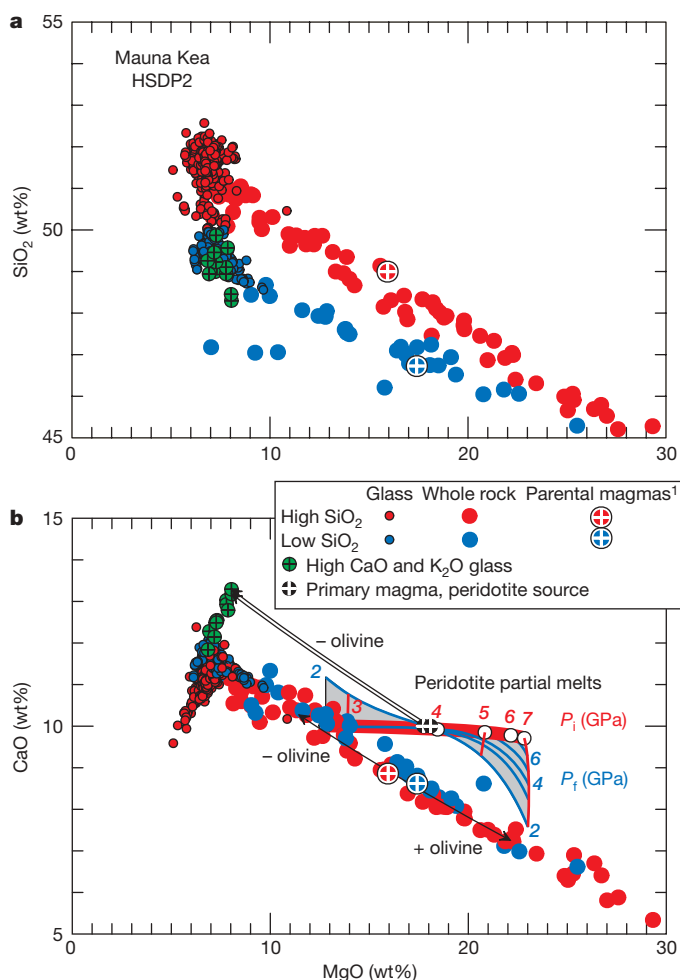


Figure 1 | Geochemical variations in Mauna Kea lavas from the Hawaii Scientific Drilling Project 2 (refs 1, 5) compared with estimated compositions of parental and primary magmas. For clarity, only whole-rock samples for which isotopic data also exist are shown. **a**, Variations in MgO and SiO₂ for glass and whole-rock samples. **b**, Variations in MgO and CaO glass and whole-rock samples compared with primary magmas of fertile mantle peridotite, formed by accumulated fractional melting (see Supplementary Information). Several glasses with high CaO and K₂O could have formed by olivine fractionation from the peridotite-source primary magmas. Parental magmas in equilibrium with Fo_{90.5} olivine¹ for commoner low- and high-SiO₂ lavas have CaO contents that are too low to qualify as partial melts and primary magmas of normal mantle peridotite. Black arrows in **b** display the effects of olivine addition (right pointing) and subtraction (left pointing) for the primary and parental magmas. Accumulated fractional melt compositions are shaded in grey, with red lines defining initial melting pressures *P_i* and blue lines defining final melting pressures *P_f*. Solidus melts are shown as white circles connected with a thick red line. Examples of how to read this diagram are given as a tutorial in the Supplementary Fig. 3 legend.

Table 2 | Estimated compositions of parental magmas and pyroxenite-source primary magmas for Mauna Kea

Type sample number	Low Si (ref. 1)	Low Si (SR0756-11.70)	Low Si (SR0823-17.50)	High Si (ref. 1)	High Si (SR0711-15.50)	High Si (SR0962-4.10)
SiO ₂	46.72	47.63	46.6	48.99	48.60	49.45
TiO ₂	1.98	1.85	0.92	1.91	2.04	1.84
Al ₂ O ₃	10.52	10.54	10.38	10.56	9.94	10.49
Cr ₂ O ₃	0	0	0	0	0	0
Fe ₂ O ₃	0	0.926	0.983	0	0.873	0.856
FeO	11.76	10.692	10.98	10.79	10.048	9.724
MnO	0.17	0	0	0.17	0	0
MgO	17.42	17.44	18.44	15.95	17.25	16.56
CaO	8.62	8.79	8.43	8.86	9.12	8.87
Na ₂ O	1.81	1.69	1.83	1.74	1.61	1.78
K ₂ O	0.29	0.28	0.28	0.3	0.29	0.28
NiO	0	0	0	0	0	0
P ₂ O ₅	0.16	0.143	0.15	0.16	0.227	0.159
<i>k_d</i>	0.320	0.306	0.301	0.320	0.310	0.313
Olivine Mg#	90.5	90.5	90.9	90.5	90.8	90.6
<i>T</i> (°C)	1,398	1,398	1,418	1,367	1,394	1,380

Some compositions were estimated by Stolper *et al.*¹. All other compositions have been estimated by addition of olivine into the sample numbers given, and the calculation was stopped when the composition reached the cotectic [L+opx+cpx+gt]. FeO/ΣFe is assumed to be 0.90 in samples before olivine addition. The Fe–Mg exchange coefficient *k_d* for olivine–L and Mg# for olivine are calculated using the method of Toplis²⁹ at 1 atmosphere. *T* (°C) is the olivine liquidus temperature calculated at 1 atmosphere using the method of Beattie³⁰.

as possible. Nevertheless, whole rocks always exhibit considerable scatter, unlike glasses. These observations indicate that Mauna Kea volcanic glasses may be faithfully recording primary pyroxenite magmas preserved at the micrometre scale. In contrast, whole rocks record considerable averaging of these magmas and do not reveal details of melting at the resolution of glasses. Similar conclusions have been made when comparing analyses of olivine-hosted melt inclusions and whole rocks^{4,15}.

Figure 4 depicts the pyroxenite-source primary magma compositions that gave rise to the lavas as sampled in the drill core. The position of the primary magmas with respect to the thermal divide is particularly revealing in SiO₂, which imposes a rough symmetry for low- and high-SiO₂ magmas. The hottest primary magmas are identified at about 2,100 m.b.s.l. in the drill core where the low- and high-SiO₂ primary magmas are most similar in SiO₂. A drop in

magmatic temperature can be inferred above and below 2,100 m.b.s.l., where SiO₂ decreases and MgO increases for the low-SiO₂ primary pyroxenite magmas. This behaviour for MgO might be counter-intuitive, but liquids change in this way when temperatures decrease for the cotectic [L+opx+cpx+gt] on the olivine-rich side of the divide. For reasons that are not clear, the glasses have somewhat lower Al₂O₃ than do the whole rocks, and this yields primary magmas with higher MgO contents. Small variations in Al₂O₃ are responsible for the variable and noisy MgO signal, but SiO₂ is not as strongly affected.

I interpret the rough stratigraphic symmetry seen in Fig. 4 to mean a constant source temperature at any specific level in the core, and temperature-induced changes with time. Temporal variations in the raw glass compositions were also discussed by Stolper *et al.*¹, but they occur at different core depths, indicating a separate part played by magma chambers or melt transport¹. The temperature variations

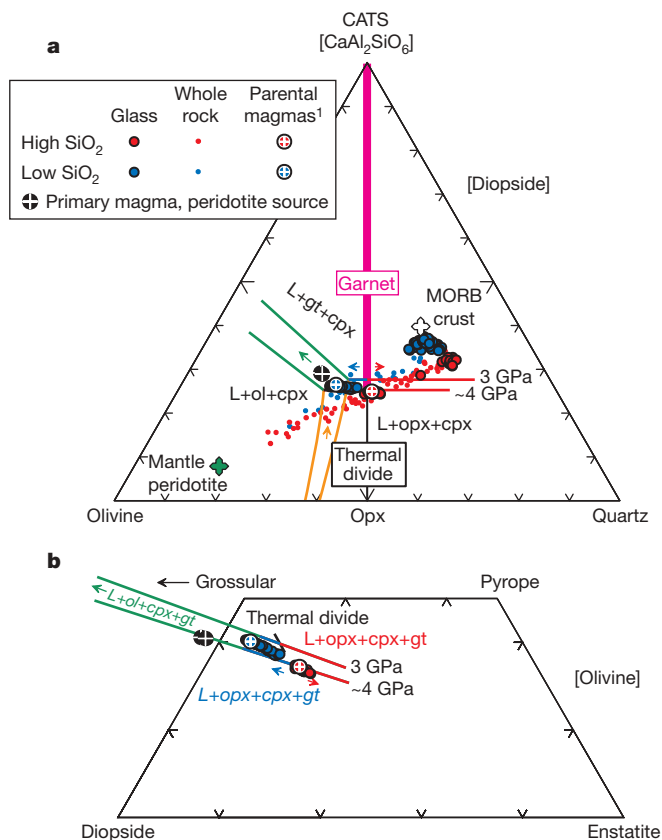


Figure 2 | Projections of HSDP2 whole rocks and glasses compared with cotectics at 3 and 4 GPa, which are arrays of liquid compositions that co-crystallize various crystalline phases. Of special interest here is the cotectic that defines liquid compositions that co-crystallize orthopyroxene, clinopyroxene and garnet [L+opx+cpx+gt], in blue and red for segments on the olivine-rich and SiO₂-rich sides of the thermal divide, respectively. Other cotectics are [L+olivine+opx±cpx], in brown, and [L+olivine+cpx+gt], in green. The projected cotectics are discussed in the Supplementary Information. Glasses have been filtered out to include only those samples that have experienced olivine fractionation (that is, >7.5% MgO). Mid-ocean-ridge basalt (MORB) crust is the experimental composition used by ref. 13. Peridotite composition is from ref. 14. **a**, A mole% projection from or towards diopside into the plane olivine–quartz–CATS. The pyroxene–garnet plane is represented by the line CATS–opx, and it is a thermal divide, as shown by opposing arrows next to the cotectic [L+opx+cpx+gt]. Shown are individual glass analyses from ref. 1 that plot furthest from olivine, and the same glass analyses into which olivine was added until the desired composition reached the cotectic [L+opx+cpx+gt]; these are primary pyroxenite magma compositions. **b**, A mole% projection from or towards olivine into part of the pyroxene–garnet plane. High-SiO₂ glasses are coincident with the cotectic [L+opx+cpx+gt] on the silica-rich side of the thermal divide in Fig. 2a, shown in red. Low-SiO₂ glasses are coincident with the cotectic [L+opx+cpx+gt] on the olivine-rich side of the thermal divide in Fig. 2a, shown in blue. The parental magmas of ref. 1 are very similar to primary magmas of pyroxenite. For whole rocks, the low- and high-SiO₂ populations generally project on each side of the thermal divide and, although this is not shown for purposes of clarity, there is considerable overlap (Supplementary Information). A three-dimensional geochemical representation is obtained by a simultaneous visualization of panels **a** and **b**. Diopside and enstatite are pyroxenes: Ca(Mg,Fe)Si₂O₆ and (Mg,Fe)₂Si₂O₆. Grossular and pyrope are garnets: Ca₃Al₂Si₃O₁₂ and (Mg,Fe)₃Al₂Si₃O₁₂.

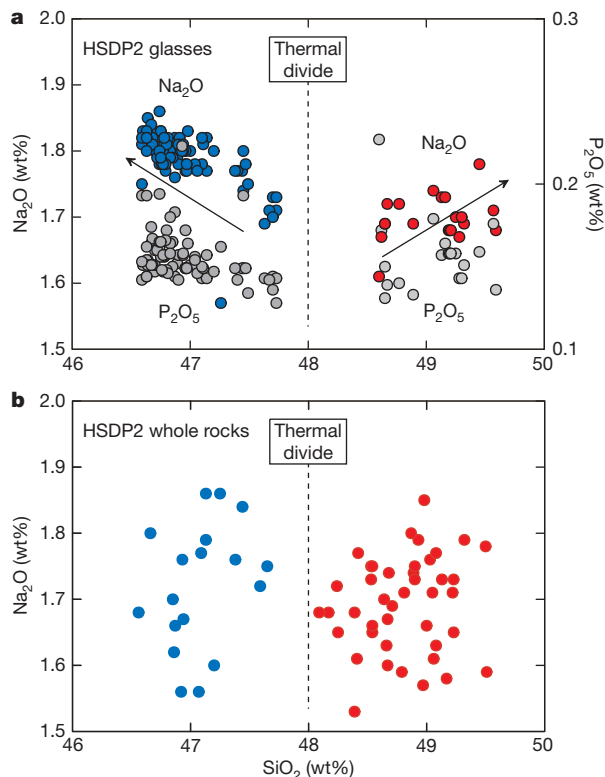


Figure 3 | Na_2O and P_2O_5 proxies for temperature along the cotectic $[\text{L} + \text{opx} + \text{cpx} + \text{gt}]$. Maximum temperatures occur at the thermal divide for primary magmas near the pyroxene–garnet plane, where liquids contain $\sim 48\%$ SiO_2 . Arrows point in the direction of decreasing temperature. For constant pyroxenite source composition, liquids near this plane are predicted to have minimum contents of Na_2O and P_2O_5 . **a**, Primary pyroxenite magma compositions determined from glass analyses. Glasses identified by ref. 1 at 2,233 m.b.s.l. are similar to low- SiO_2 types, but they are highest in Na_2O and MgO , and plot at the lowest temperatures in this group. **b**, Primary pyroxenite magma compositions determined from whole-rock analyses filtered to exclude samples for which there has been minor plagioclase and augite addition and subtraction (see Supplementary Information).

displayed by the low- SiO_2 primary pyroxenite magmas are expected to be $30\text{--}50^\circ\text{C}$, based on the following temperatures at 3 GPa: $1,560\text{--}1,580^\circ\text{C}$ near the thermal divide^{16,17}, and $1,515\text{--}1,540^\circ\text{C}$ at the olivine termination of the cotectic $[\text{L} + \text{ol} + \text{opx} + \text{cpx} + \text{gt}]$ ^{3,8}. Peridotite-source melts appear briefly at about 1,800 m.b.s.l., and these will have a 3 GPa liquidus temperature of about $1,550^\circ\text{C}$ (ref. 8) and a potential temperature that is also $1,550^\circ\text{C}$; initial and final melting pressures are about 4.0 and 3.0 GPa, respectively. Pyroxenite source temperature variations translate to potential temperatures of $\sim 1,500\text{--}1,550^\circ\text{C}$. This is evidence for magmatic sampling of different temperature domains within the plume, the axis being the hottest.

Temperatures are $1,470\text{--}1,500^\circ\text{C}$ at the 3 GPa anhydrous peridotite solidus^{18,19} and $1,560\text{--}1,580^\circ\text{C}$ near the thermal divide^{16,17}. It is therefore possible for pyroxenite melts with compositions along the cotectic $[\text{L} + \text{opx} + \text{cpx} + \text{gt}]$ and near the thermal divide to be hotter than peridotite melts even though they are lower in MgO . The common petrological practice of using MgO in a primary magma to infer volcanic and source temperature conditions should be made judiciously, and is only secure when the source is olivine-bearing.

The Hawaiian tholeiites cannot be single-stage partial melts of an original basaltic crustal protolith of biminerally eclogite or quartz/coesite eclogite^{13,14} at 3.0–3.5 GPa because they are too high in NiO and MgO , and too low^{2,13} in SiO_2 and Al_2O_3 (Fig. 2). Sobolev *et al.*² proposed that the pyroxenite source forms in a second stage by melt-rock reaction. At 3 GPa, quartz eclogite begins to melt at about $1,315^\circ\text{C}$ ¹³ and the SiO_2 -rich melts can react with the peridotite host to produce^{14,20} $\text{opx} + \text{cpx} + \text{gt}$ (Supplementary Information). Orthopyroxene may be heavily concentrated in bands, as indicated by experiment^{14,20} and metasomatic theory^{2,21}, but primary magmas will form at contacts with cpx and gt where the temperatures are at a minimum on the cotectic $[\text{L} + \text{opx} + \text{cpx} + \text{gt}]$. Large melt fractions of pyroxenite may be temporarily entombed by these orthopyroxene-rich reaction zones²², but leaks might refertilize the surrounding peridotite.

The phase diagram in Fig. 2 provides a key for a general understanding of melt production in lithologically heterogeneous mantle. Unlike pyroxenite melting below other oceanic islands on Earth, pyroxenite melts with compositions on the SiO_2 -rich side of the thermal divide are unique to the shield-building lavas of Hawaii. And, unlike Mauna Kea, there is no evidence for the involvement of peridotite-source melts below Mauna Loa and Kilauea. For all shield-building lavas on Hawaii, the phase diagram requires a substantial role for SiO_2 -rich basaltic oceanic crust (Fig. 2a). This is

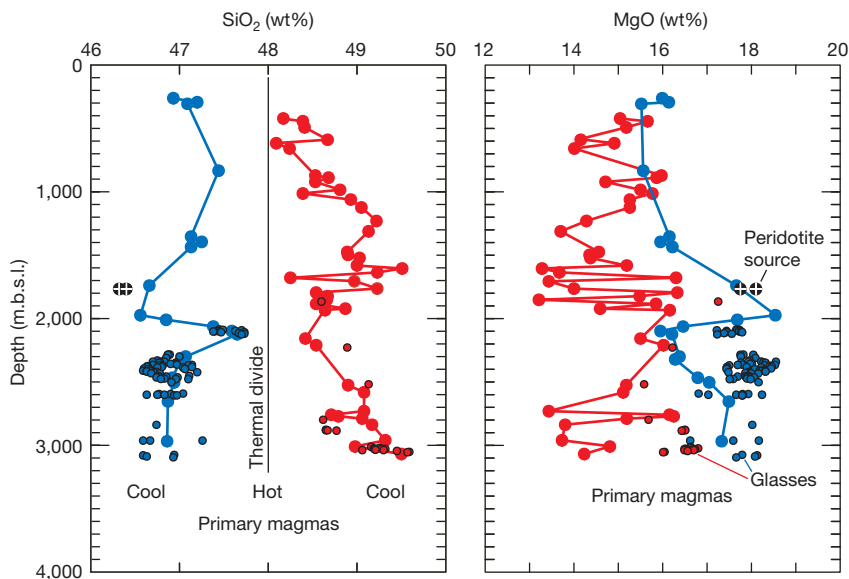


Figure 4 | Variations in SiO_2 and MgO contents of pyroxenite primary magma compositions for glasses and whole rocks with depth in the HSDP2 drill core. Whole-rock analyses have been filtered to exclude samples for which there has been minor plagioclase and augite addition and subtraction; these are connected by the blue (SiO_2 -poor) and red (SiO_2 -rich) lines. Glasses have been filtered to include only those samples that have experienced olivine fractionation (that is, $>7.5\%$ MgO). Core locations are shown in metres below sea level (m.b.s.l.). SiO_2 contents of pyroxenite primary magmas are good proxies for temperatures of their source, with maximum temperatures realized for magmas with $\sim 48\%$ SiO_2 produced nearest the thermal divide.

supporting evidence for previous suggestions that recycled crust is organized in large bodies^{23,24}, reconstituted as pyroxenite². The implication is that oceanic crust has been subducted, stirred, stretched and returned in a plume with its fine structure roughly preserved as geochemical heterogeneities in Hawaiian volcanoes^{15,25–28}. This improbable idea seems quite reasonable from a petrological point of view.

Received 2 March; accepted 13 September 2006.

1. Stolper, E., Sherman, S., Garcia, M., Baker, M. & Seaman, C. Glass in the submarine section of the HSDP2 drill core, Hilo, Hawaii. *Geochem. Geophys. Geosyst.* 5, doi:10.1029/2003GC000553 (2004).
2. Sobolev, A. V., Hofmann, A. W., Sobolev, S. V. & Nikogosian, I. K. An olivine-free mantle source of Hawaiian shield basalts. *Nature* 434, 590–597 (2005).
3. Walter, M. J. Melting of garnet peridotite and the origin of komatiite and depleted lithosphere. *J. Petrol.* 39, 29–60 (1998).
4. Ren, Z.-Y., Ingle, S., Takahashi, E., Hirano, N. & Hirata, T. The chemical structure of the Hawaiian plume. *Nature* 436, 837–840 (2005).
5. Rhodes, J. M. & Vollinger, M. J. Composition of basaltic lavas sampled by phase-2 of the Hawaii Scientific Drilling Project: geochemical stratigraphy and magma types. *Geochem. Geophys. Geosyst.* 5, doi:10.1029/2002GC000434 (2004).
6. Frey, F. A. *et al.* The evolution of Mauna Kea volcano, Hawaii: petrogenesis of tholeiitic and alkalic basalts. *J. Geophys. Res. B* 96, 14347–14375 (1991).
7. Sobolev, A. V. & Nikogosian, I. K. Petrology of long-lived mantle plume magmatism: Hawaii, Pacific, and Reunion islands, Indian Ocean. *Petrology* 2, 111–144 (1994).
8. Herzberg, C. & O'Hara, M. J. Plume-associated ultramafic magmas of Phanerozoic age. *J. Petrol.* 43, 1857–1883 (2002).
9. Arndt, N. T., Kerr, A. C. & Tarney, J. Dynamic melting in plume heads: the formation of Gorgona komatiites and basalts. *Earth Planet. Sci. Lett.* 146, 289–301 (1997).
10. O'Hara, M. J. & Yoder, H. S. Jr. Formation and fractionation of basic magmas at high pressures. *Scott. J. Geol.* 3, 67–117 (1967).
11. Milholland, C. S. & Presnall, D. C. Liquidus phase relations in the CaO–MgO–Al₂O₃–SiO₂ system at 3.0 GPa: The aluminous pyroxene thermal divide and high pressure fractionation of picritic and komatiitic magmas. *J. Petrol.* 39, 3–27 (1998).
12. Kogiso, T., Hirschmann, M. M. & Pertermann, M. High-pressure partial melting of mafic lithologies in the mantle. *J. Petrol.* 45, 2407–2422 (2004).
13. Pertermann, M. & Hirschmann, M. M. Anhydrous partial melting experiments on MORB-like eclogite: phase relations, phase compositions and mineral-melt partitioning of major elements at 2–3 GPa. *J. Petrol.* 44, 2173–2201 (2003).
14. Yaxley, G. M. & Green, D. H. Reactions between eclogite and peridotite: mantle refertilisation by subduction of oceanic crust. *Schweiz. Mineral. Petrogr. Mitt.* 78, 243–255 (1998).
15. Sobolev, A. V., Hofmann, A. W. & Nikogosian, I. K. Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lavas. *Nature* 404, 986–989 (2000).
16. Ito, K. & Kennedy, G. C. Melting and phase relations in the plane tholeiite-lherzolite-nepheline basanite to 40 kilobars and geological implications. *Contrib. Mineral. Petrol.* 19, 177–211 (1968).
17. Maaloe, S. The PT-phase relations of an MgO-rich Hawaiian tholeiite: the compositions of primary Hawaiian tholeiites. *Contrib. Mineral. Petrol.* 148, 236–246 (2004).
18. Hirschmann, M. M. Mantle solidus: experimental constraints and the effects of peridotite composition. *Geochem. Geophys. Geosyst.* 1, doi:10.1029/2000GC000070 (2000).
19. Herzberg, C., Ratteron, P. & Zhang, J. New experimental observations on the anhydrous solidus for peridotite KLB-1. *Geochem. Geophys. Geosyst.* 1, doi:10.1029/2000GC000089 (2000).
20. Takahashi, E. & Nakajima, K. in *Hawaiian Volcanoes: Deep Underwater Perspectives* (eds Takahashi, E., Lipman, P. W., Garcian, M. O., Naka, J. & Aramaki, S.) Vol. 128, 403–418 (Geophys. Monogr. Ser., AGU, Washington DC, 2002).
21. Korzhinskii, D. S. *Theory of Metasomatic Zoning* (Clarendon, Oxford, 1970).
22. Kogiso, T., Hirschmann, M. M. & Reiners, P. W. Length scales of mantle heterogeneities and their relationship to ocean island basalt geochemistry. *Geochim. Cosmochim. Acta* 68, 345–360 (2004).
23. Abouchami, W. *et al.* Lead isotopes reveal bilateral asymmetry and vertical continuity in the Hawaiian mantle plume. *Nature* 434, 851–856 (2005).
24. Blichert-Toft, J., Weis, D., Maerschalk, C., Agrani, A. & Albarède, F. Hawaiian hot spot dynamics as inferred from the Hf and Pb isotope evolution of Mauna Kea volcano. *Geochem. Geophys. Geosyst.* 4, doi:10.1029/2002GC000340 (2003).
25. Lassiter, J. C. & Hauri, E. H. Osmium-isotope variations in Hawaiian lavas: evidence for recycled lithosphere in the Hawaiian plume. *Earth Planet. Sci. Lett.* 164, 483–496 (1998).
26. Blichert-Toft, J., Frey, F. A. & Albarède, F. Hf isotope evidence for pelagic sediments in the source of Hawaiian basalts. *Science* 285, 879–882 (1999).
27. Huang, S. *et al.* Enriched components in the Hawaiian plume: Evidence from Kahoolawe volcano, Hawaii. *Geochem. Geophys. Geosyst.* 6, doi:10.1029/2005GC001012 (2005).
28. Hauri, E. Major element variability in the Hawaiian plume. *Nature* 382, 415–419 (1996).
29. Toplis, M. J. The thermodynamics of iron and magnesium partitioning between olivine and liquid: criteria for assessing and predicting equilibrium in natural and experimental systems. *Contrib. Mineral. Petrol.* 149, 22–39 (2005).
30. Beattie, P. Olivine-melt and orthopyroxene-melt equilibria. *Contrib. Mineral. Petrol.* 115, 103–111 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I am grateful to A. Sobolev, M. Feigenson, M. Hirschmann, F. Frey and P. Asimow for discussions. I also thank A. Hofmann and A. Sobolev for critical comments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.H. (herzberg@rci.rutgers.edu).

LETTERS

Boosting slow oscillations during sleep potentiates memory

Lisa Marshall¹, Halla Helgadóttir¹, Matthias Mölle¹ & Jan Born¹

There is compelling evidence that sleep contributes to the long-term consolidation of new memories¹. This function of sleep has been linked to slow (<1 Hz) potential oscillations, which predominantly arise from the prefrontal neocortex and characterize slow wave sleep^{2–4}. However, oscillations in brain potentials are commonly considered to be mere epiphenomena that reflect synchronized activity arising from neuronal networks, which links the membrane and synaptic processes of these neurons in time⁵. Whether brain potentials and their extracellular equivalent have any physiological meaning *per se* is unclear, but can easily be investigated by inducing the extracellular oscillating potential fields of interest^{6–8}. Here we show that inducing slow oscillation-like potential fields by transcranial application of oscillating potentials (0.75 Hz) during early nocturnal non-rapid-eye-movement sleep, that is, a period of emerging slow wave sleep, enhances the retention of hippocampus-dependent declarative memories in healthy humans. The slowly oscillating potential stimulation induced an immediate increase in slow wave sleep, endogenous cortical slow oscillations and slow spindle activity in the frontal cortex. Brain stimulation with oscillations at 5 Hz—another frequency band that normally predominates during rapid-eye-movement sleep—decreased slow oscillations and left declarative memory unchanged. Our findings indicate that endogenous slow potential oscillations have a causal role in the sleep-associated consolidation of memory, and that this role is enhanced by field effects in cortical extracellular space.

Slow oscillations reflect widespread ‘up’ and ‘down’ states of network activity. These oscillations are generated within the neocortex and are most prominent during slow wave sleep (SWS); the up and down states reflect, respectively, global neuronal depolarization with excitation, and neuronal hyperpolarization with neuronal silence^{2,3,9}. Essentially owing to its synchronizing influence on neuronal activity within the neocortex and in dialogue with thalamic and hippocampal circuitry, the slow oscillation has been suspected to underlie the consolidation of memory during sleep^{3,10–13}. The slow oscillation signal peaks at 0.7–0.8 Hz, although spectral components can extend into the slow delta band (1–4 Hz)^{9,14}. We have examined the role of slow oscillations in memory consolidation by inducing them through transcranial application of oscillating potentials during early nocturnal non-rapid-eye-movement (non-REM) sleep after a learning period.

Hippocampus-dependent declarative memory was assessed by a paired-associate learning task, with memory retention measured by the difference in the number of words recalled when tested after sleep and the number recalled at learning before sleep. As expected from previous studies using the same procedure^{8,15}, retrieval testing after sleep showed an increase in performance, compared to learning before sleep, in both the slow oscillation stimulation and the sham stimulation conditions (Fig. 1b). However, after slow oscillation

stimulation this increase in memory (mean 4.77 words) was greater than that following sham stimulation (mean 2.08 words; $F_{1,12} = 7.96$, $P = 0.01$). In the stimulation condition, 36.50 ± 1.24 words were recalled at learning before sleep, and this number increased to 41.27 ± 1.21 at retrieval after sleep. In the sham condition performance increased from 37.42 ± 0.92 (learning) to 39.50 ± 0.84 words after sleep. This improvement in retention following stimulation is striking considering that most subjects were medical students, who were highly trained in memorizing facts and already performed well in the sham condition. Note that our retention measure does not allow us to differentiate between a slowed decay and an actual gain in memory after stimulation¹⁶.

To test whether slow oscillation stimulation specifically affected the formation of hippocampus-dependent declarative memory, we also tested subjects on a non-declarative, procedural finger-sequence tapping task¹⁷. Retrieval testing after sleep confirmed the characteristic overnight improvement in skill in both conditions (number of

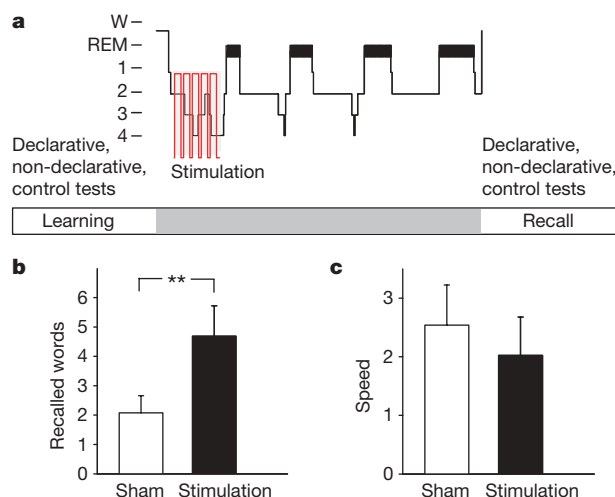


Figure 1 | Slow oscillatory stimulation enhances declarative memory performance. **a**, Time-course of experiment. Indicated are time points of learning and recall of memory tasks, psychometric control tests, stimulation intervals, period of lights off (horizontal grey bar), and sleep represented by a hypnogram. W, wake; 1–4, sleep stages 1–4. **b**, Performance on the declarative paired-associate memory task across the retention period of nocturnal sleep for stimulation and sham stimulation. Performance is expressed as difference between the number of correct words reported at recall testing and learning. The list contained 46 experimental word-pairs (** $P < 0.01$). **c**, Performance speed on the non-declarative procedural motor skill task across the retention interval expressed as the difference in the number of correctly tapped sequences per 30 s between recall testing and learning. Data are the means $\pm$ s.e.m.

¹University of Lübeck, Department of Neuroendocrinology, Haus 23a, Ratzeburger Allee 160, 23538 Lübeck, Germany.

Table 1 | Sleep parameters during transcranial stimulation and sham conditions

Slow oscillation stimulation (<i>n</i> = 13)	Stimulation (mean $\pm$ SEM)	Sham (mean $\pm$ SEM)	Theta stimulation (<i>n</i> = 5)	Stimulation (mean $\pm$ SEM)	Sham (mean $\pm$ SEM)
Awake	0.8 $\pm$ 0.8	1.5 $\pm$ 1.5	Awake	12.0 $\pm$ 12.0	0.0 $\pm$ 0.0
S1	17.7 $\pm$ 10.3	30.8 $\pm$ 12.1	S1	18.2 $\pm$ 18.2	24.1 $\pm$ 24.1
S2	110.8 $\pm$ 14.2	143.1 $\pm$ 20.4	S2	181.6 $\pm$ 38.0	144.1 $\pm$ 30.6
S3	113.1 $\pm$ 18.3	85.4 $\pm$ 16.2	S3	59.3 $\pm$ 34.2	108.1 $\pm$ 24.4*
S4	57.7 $\pm$ 17.1	39.2 $\pm$ 18.7	S4	29.8 $\pm$ 13.4	23.9 $\pm$ 17.5
SWS	170.8 $\pm$ 17.8	124.6 $\pm$ 19.2*	SWS	89.1 $\pm$ 42.4	132.0 $\pm$ 35.0

Time (s) spent in different sleep stages during the 1-min stimulation-free intervals between the periods of stimulation for the main experiment (top) testing slow oscillation stimulation (0.75 Hz) and in a supplementary experiment (bottom) testing theta stimulation (5 Hz). Sleep scoring based on 10-s intervals. Awake and S1 sometimes occurred in just one case. Asterisk indicates a significant ($P < 0.05$) increase in SWS with slow oscillation stimulation and a significant ($P < 0.05$, one-tailed) decrease in S3 sleep with theta stimulation. S1–S4, sleep stages 1–4.

correctly tapped sequences before sleep 17.74 ± 1.30 (stimulation), 18.15 ± 1.28 (sham), and at retrieval after sleep 19.77 ± 1.52 and 20.69 ± 1.46 , respectively; $F_{1,12} = 67.70$, $P < 0.001$). However, in contrast to declarative memory performance, the sleep-associated gain in performance (2.03 ± 0.65 sequences) was not enhanced through slow oscillation stimulation ($P > 0.6$; Fig. 1c). Also, overnight changes in error rate did not differ between the two conditions ($P > 0.25$). Performance on two additional tasks, a declarative non-verbal paired-associate task and a procedural mirror-tracing task, likewise indicated that stimulation led to an improvement only for the declarative task (see Supplementary Information for a figure summarizing the main result). The ability of slow oscillation stimulation during early non-REM sleep to enhance retention of word-pairs and its failure to affect procedural skill are consistent with reports that hippocampus-dependent memories benefit mainly from early SWS, and procedural memories from REM sleep (which prevails during late sleep), although non-REM sleep might have complementary functions^{1,18,19}. The efficacy of polarization over the prefrontal cortex in our study is in line with this region's importance

in the hippocampal–neocortical dialogue that is assumed to underlie the consolidation of hippocampus-dependent memories²⁰.

In a control experiment ($n = 8$) using a protocol identical to that of the main experiment we shifted the timing of stimulation to the period shortly before awakening (05.45–06.15 h), that is closer to retrieval testing, which should increase any immediate non-specific effects of stimulation on cognitive function during retrieval²¹. However, under this condition retention of word-pairs remained unchanged and was similar to retention after sham stimulation (post-sleep retrieval with reference to learning: 3.21 ± 1.43 versus 2.93 ± 1.21 words, $P > 0.8$). These and further control tests of vigilance and general retrieval capabilities (see Supplementary Information) safely exclude any substantial non-specific contribution of slow oscillation stimulation to the improved declarative memory at retrieval testing.

We examined sleep and electroencephalogram (EEG) activity more closely to understand the mechanisms that underlie the enhancement of memory performance. During the 5-min periods of acute stimulation, the induced potentials precluded sleep scoring (Fig. 2). However, the 1-min stimulation-free intervals yielded clear signals. During these intervals, more total time was spent in SWS after slow oscillation (170.77 ± 17.78 s) than after sham stimulation (124.62 ± 19.17 s, $F_{1,12} = 6.03$, $P < 0.05$; Table 1). The times spent in the different sleep stages during the 60 minutes after stimulation and for the whole night were comparable between stimulation and sham conditions (Supplementary Table 1). The increase in SWS is a plausible explanation for the ability of slow oscillatory stimulation to improve memory, particularly as this increase was presumably also present during the periods of acute stimulation. However, visual scoring of SWS relies strongly on an undifferentiated estimation of the presence of slow wave rhythms. We suspected that the transient increase in SWS during the 1-min intervals between periods of stimulation reflected a temporally limited enhancement of EEG slow oscillations, and that this was the specific process that, during this sleep stage, mediated the memory improvement.

Indeed, spectral analysis of EEG activity during the five 1-min intervals between stimulation periods confirmed that stimulation acutely facilitated endogenous slow oscillations. Stimulation distinctly enhanced EEG power within the slow oscillation band (0.5–1.0 Hz, $F_{1,12} = 11.67$, $P < 0.01$ at the frontocentral recording site, Fz; Fig. 3a). A slight increase in power in adjacent low (1–1.5 Hz) delta frequencies and in the 1–4-Hz delta band failed to reach significance ($P > 0.06$, for all comparisons) indicating that the effect of stimulation was focused on the slow oscillation band. Interestingly, slow oscillation stimulation simultaneously enhanced EEG power within the slow spindle frequency range (8–12 Hz, peaking at ~ 10.5 Hz, $F_{1,12} = 13.12$, $P < 0.01$ at Fz; Fig. 3a) as well as spindle counts (see Supplementary Information). The effects were most pronounced for the first three inter-stimulation intervals (Fig. 3b), with a prefrontal maximum, although they spread to the other recording sites ($F_{1,12} > 7.43$, $P < 0.02$, for overall effects of stimulation). The conjunct increase in slow oscillation and frontal spindle activity agrees well with the notion that neocortical slow oscillations drive the thalamic generation of spindles^{3,9,14} and emphasizes that stimulation induces a physiologically coherent pattern of activity in this system.

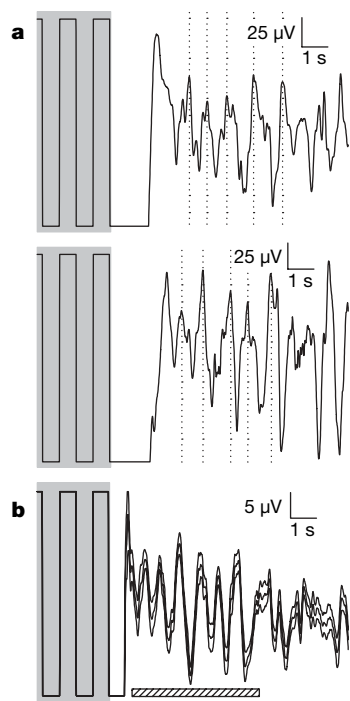


Figure 2 | Synchronization of slow oscillatory EEG activity. **a**, EEG recordings during the last seconds of a 5-min stimulation period (shaded areas) and first few seconds of a stimulation-free interval of two individuals at prefrontal sites (Fz). **b**, Corresponding mean $\pm$ s.e.m. across all subjects and stimulation periods over the parietal cortex (where the EEG is least contaminated by the ceasing stimulation artefact). Positivity upward. Note entrainment of the slow oscillatory EEG activity to the slow oscillatory rhythmic stimulation. Hatched bar indicates time interval of stimulation-induced phase changes in the 0.78–0.98-Hz and 1.37–1.56-Hz bins of the EEG signal.

Inspection of recordings in some individuals clearly revealed the presence of slow oscillations at the same frequency and in phase with those induced by stimulation, right after stimulation ceased, indicating that the cortical activity boosted by slow oscillation stimulation continued into the stimulation-free period (Fig. 2). In fact, the EEG phase distribution, for a 5-s interval after stimulation ceased, was changed compared to sham stimulation specifically for the frequency bin that approximated the stimulation frequency as well as for the bin that corresponded to twice the stimulation frequency ($P < 0.05$, Kolmogorov-Smirnov; bins 0.78–0.98 Hz and 1.37–1.56 Hz). It has been shown *in vitro* that neuronal networks can synchronize to oscillating electrical fields, showing resonance around particular frequencies^{7,22}.

We tested the view that resonance specifically in the slow oscillation band is essential for the beneficial effect on memory by using stimulation oscillating at 5 Hz (corresponding to the physiological theta rhythm) under conditions otherwise identical to the main experiment. Compared with sham stimulation, theta stimulation did not increase but rather reduced slow oscillation power in all subjects ($n = 5$) although to variable degrees (at Fz averaged across the first three stimulation-free 1-min intervals: 33.8 ± 14.7 versus $71.5 \pm 20.8 \mu\text{V}^2$ in the sham condition, $P < 0.045$, Wilcoxon). Likewise, theta stimulation in these intervals did not increase, but

rather non-significantly decreased slow frontal spindle power (0.60 ± 0.12 versus $0.78 \pm 0.14 \mu\text{V}^2$, $P = 0.34$). Sleep during theta stimulation contained less stage 3 SWS than during sham stimulation (Table 1). Finally, theta stimulation also did not improve declarative memory for word-pairs (post-sleep retrieval with reference to learning: 2.2 ± 1.4 versus 2.4 ± 0.75 words after sham stimulation, $P > 0.5$).

Our results indicate that slow oscillations have a causal role in consolidating hippocampus-dependent memories during sleep. How could slow oscillations promote the plastic neuronal changes that underlie such memory consolidation? One plausible mechanism might involve calcium transients mediated by spindle activity^{2,23,24}, as spindle activity was enhanced by slow oscillation stimulation. Not only is spindle activity probably associated with massive Ca^{2+} influx into neocortical pyramidal cells, but there is also evidence that repeated spindle-associated spike discharges can trigger long-term potentiation in neocortical synapses²⁵. As synchronous spindle activity occurs preferentially at synapses previously potentiated by tetanizing afferent stimulation²⁶, slow-oscillation-driven spindle activity might contribute to the strengthening of synaptic connections in neocortical circuitry.

Notably, our stimulation induced an estimated potential field in extracellular space that closely resembled that accompanying endogenous slow oscillations. The electric field at the cortical surface right below our stimulation electrode reached an estimated amplitude of $\sim 1,200 \mu\text{V mm}^{-1}$ (refs 27, 28). Calculations from local field potential recordings *in vivo* during slow oscillatory activity similarly indicate potential fields up to $1,600 \mu\text{V mm}^{-1}$, depending on the distance to the slow oscillation generator (see, for example, ref. 29). Neurons can synchronize to much weaker oscillating fields, as small as $295 \mu\text{V mm}^{-1}$ (ref. 22). The synchronizing effect is stronger for neuronal networks than for single neurons, and persists in the absence of functioning chemical synapses⁶. On this background, our results challenge the common view that extracellular slow potential oscillations represent mere epiphenomena without physiological significance *per se*. Because of the insulating property of neuronal membranes, the transcranial stimulation reaches the extracellular space first. Our data therefore indicate that an extracellular slow potential oscillation comparable with that accompanying brain-borne slow oscillations is itself sufficient to increase memory, implicating field effects in cognitive processing during sleep.

METHODS

Oscillating potential fields were induced in young healthy humans ($n = 13$) through stimulation electrodes applied bilaterally at frontolateral locations and at the mastoids, with the frontolateral electrodes representing sites of anodal (positive) polarization. We induced potentials by transcranially applying currents oscillating at a frequency of 0.75 Hz (maximum current density: 0.517 mA cm^{-2}). Stimulation started 4 min after subjects had entered non-REM sleep stage 2 for the first time (without transitions back to stage 1 sleep or wakefulness), that is a time when sleep is expected to progress into SWS. It was applied for five 5-min intervals separated by 1-min intervals free of stimulation. Subjects were tested twice, in a stimulation condition and a sham stimulation condition. In each condition, in the evening before sleep (21.00–22.30 h), subjects learned to a criterion different memory tasks. Recall of memories was tested the following morning (7.00–7.30 h; Fig. 1a). Declarative memory was assessed by a paired-associate learning task^{8,15} consisting of a list of 46 word-pairs to be learned before sleep to a criterion of 60% correct responses in a test of immediate cued recall. At retrieval testing in the morning after sleep, cue words were again displayed and the subjects were required to recall the appropriate response word. The procedural task, which has likewise proved sensitive to the enhancing effects of sleep in previous studies¹⁷, required subjects to repeatedly finger-tap with the non-dominant left hand a five-element sequence as fast and accurately as possible on a keyboard for twelve 30-s periods at learning before sleep, and for three 30-s periods at post-sleep retesting. EEG and standard polysomnography were recorded continuously during the sleep period lasting from 23.00 h (lights off) till 6.30 h (awakening; see Supplementary Information for detailed descriptions of the procedure, experimental and analytical techniques).

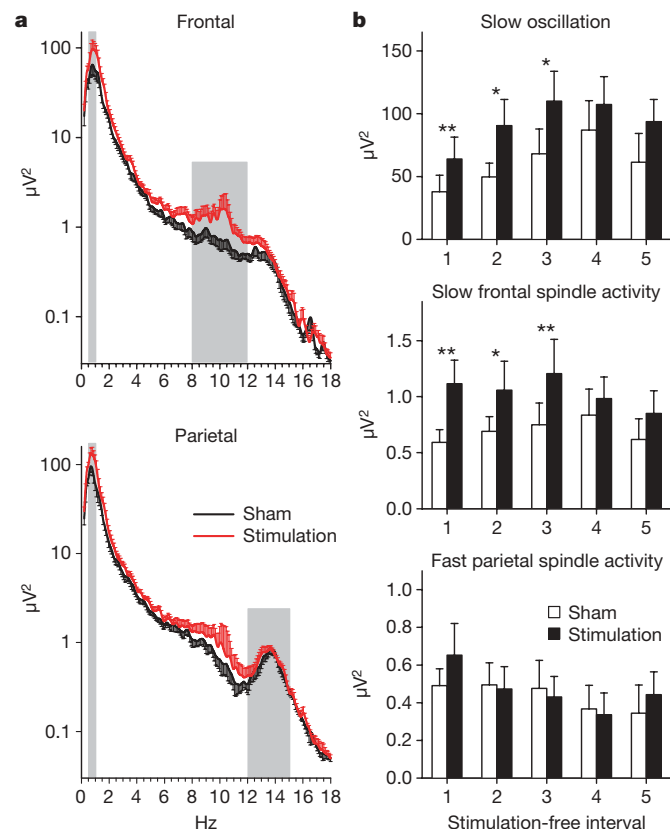


Figure 3 | EEG activity during the 1-min intervals between periods of slow oscillation stimulation and between corresponding periods of sham stimulation. **a**, Average power spectrum (across first three stimulation-free intervals) at the midline frontal and parietal sites. Shaded areas indicate frequency bands for slow oscillations (0.5–1 Hz), slow frontal spindle activity (upper panel, 8–12 Hz), and fast parietal spindle activity (lower panel, 12–15 Hz). **b**, Time course of power in the five stimulation-free intervals for slow oscillations, slow frontal spindle activity and fast parietal spindle activity. Slow frontal spindle activity is to some extent also visible over the parietal cortex, reflecting the more widespread neuronal synchrony underlying this spindle class³⁰. Stimulation enhances slow oscillation and slow spindle activity at the frontal location, but not fast spindle activity at the parietal location. Asterisks indicate statistical significance (** $P < 0.01$, * $P < 0.05$) for pairwise comparison. Data are the means $\pm$ s.e.m.

Received 19 July; accepted 25 September 2006.

Published online 5 November 2006.

1. Stickgold, R. Sleep-dependent memory consolidation. *Nature* **437**, 1272–1278 (2005).
2. Sejnowski, T. J. & Destexhe, A. Why do we sleep? *Brain Res.* **886**, 208–223 (2000).
3. Steriade, M. & Timofeev, I. Neuronal plasticity in thalamocortical networks during sleep and waking oscillations. *Neuron* **37**, 563–576 (2003).
4. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. Local sleep and learning. *Nature* **430**, 78–81 (2004).
5. Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
6. Jefferys, J. G. & Haas, H. L. Synchronized bursting of CA1 hippocampal pyramidal cells in the absence of synaptic transmission. *Nature* **300**, 448–450 (1982).
7. Hutcheon, B. & Yarom, Y. Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci.* **23**, 216–222 (2000).
8. Marshall, L., Mölle, M., Hallschmid, M. & Born, J. Transcranial direct current stimulation during sleep improves declarative memory. *J. Neurosci.* **24**, 9985–9992 (2004).
9. Steriade, M. The corticothalamic system in sleep. *Front. Biosci.* **8**, d878–d899 (2003).
10. Buzsáki, G. Memory consolidation during sleep: a neurophysiological perspective. *J. Sleep Res.* **7** (Suppl. 1), 17–23 (1998).
11. Mölle, M., Marshall, L., Gais, S. & Born, J. Learning increases human electroencephalographic coherence during subsequent slow sleep oscillations. *Proc. Natl Acad. Sci. USA* **101**, 13963–13968 (2004).
12. Mölle, M., Yeshenko, O., Marshall, L., Sara, S. J. & Born, J. Hippocampal sharp wave-ripples linked to slow oscillations in rat slow-wave sleep. *J. Neurophysiol.* **96**, 62–70 (2006).
13. Wolansky, T., Clement, E. A., Peters, S. R., Palczak, M. A. & Dickson, C. T. Hippocampal slow oscillation: a novel EEG state and its coordination with ongoing neocortical activity. *J. Neurosci.* **26**, 6213–6229 (2006).
14. Mölle, M., Marshall, L., Gais, S. & Born, J. Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *J. Neurosci.* **22**, 10941–10947 (2002).
15. Plihal, W. & Born, J. Effects of early and late nocturnal sleep on declarative and procedural memory. *J. Cogn. Neurosci.* **9**, 534–547 (1997).
16. Walker, M. P. & Stickgold, R. Sleep-dependent learning and memory consolidation. *Neuron* **44**, 121–133 (2004).
17. Walker, M. P., Brakefield, T., Hobson, J. A. & Stickgold, R. Dissociable stages of human memory consolidation and reconsolidation. *Nature* **425**, 616–620 (2003).
18. Gais, S. & Born, J. Declarative memory consolidation: mechanisms acting during human sleep. *Learn. Mem.* **11**, 679–685 (2004).
19. Peigneux, P. *et al.* Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* **44**, 535–545 (2004).
20. Wiltgen, B. J., Brown, R. A., Talton, L. E. & Silva, A. J. New circuits for old memories: the role of the neocortex in consolidation. *Neuron* **44**, 101–108 (2004).
21. Nitsche, M. A. & Paulus, W. Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* **57**, 1899–1901 (2001).
22. Francis, J. T., Gluckman, B. J. & Schiff, S. J. Sensitivity of neurons to weak electric fields. *J. Neurosci.* **23**, 7255–7261 (2003).
23. Massimini, M. & Amzica, F. Extracellular calcium fluctuations and intracellular potentials in the cortex during the slow sleep oscillation. *J. Neurophysiol.* **85**, 1346–1350 (2001).
24. Shu, Y., Hasenstaub, A., Duque, A., Yu, Y. & McCormick, D. A. Modulation of intracortical synaptic potentials by presynaptic somatic membrane potential. *Nature* **441**, 761–765 (2006).
25. Rosanova, M. & Ulrich, D. Pattern-specific associative long-term potentiation induced by a sleep spindle-related spike train. *J. Neurosci.* **25**, 9398–9405 (2005).
26. Werk, C. M., Harbour, V. L. & Chapman, C. A. Induction of long-term potentiation leads to increased reliability of evoked neocortical spindles *in vivo*. *Neuroscience* **131**, 793–800 (2005).
27. Lutzenberger, W. & Elbert, T. *Safety Assessment of NMR Clinical Equipment* (ed. Schmidt, K. H.) 36–45 (Thieme, Stuttgart, 1987).
28. Rush, S. & Driscoll, D. A. Current distribution in the brain from surface electrodes. *Anesth. Analg.* **47**, 717–723 (1968).
29. Steriade, M., Contreras, D., Amzica, F. & Timofeev, I. Synchronization of fast (30–40 Hz) spontaneous oscillations in intrathalamic and thalamocortical networks. *J. Neurosci.* **16**, 2788–2808 (1996).
30. Doran, S. M. The dynamic topography of individual sleep spindles. *Sleep Res. Online* **5**, 133–139 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A figure summarizing the main result of this paper is available in Supplementary Information.

Acknowledgements We thank H. Koller for help in designing the stimulation apparatus, H. Schuster, H. Siebner, B. Rasch and U. Wagner for discussions of our results, and A. Otterbein, S. Uyanik, P. Paul, R. Krebs and M. Rohwer for technical assistance. This work is supported by the Deutsche Forschungsgemeinschaft.

Author Contributions L.M. and H.H. conducted the experiments. L.M., M.M. and J.B. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.M. (marshall@kfg.uni-luebeck.de) or J.B. (born@kfg.uni-luebeck.de).

LETTERS

Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex

Patrick McDonel^{1,2*}, Judith Jans^{1,2*}, Brant K. Peterson² & Barbara J. Meyer^{1,2}

Gene expression in metazoans is regulated not only at the level of individual genes but also in a coordinated manner across large chromosomal domains (for example centromeres, telomeres and imprinted gene clusters^{1–3}) and along entire chromosomes (for example X-chromosome dosage compensation^{4–6}). The primary DNA sequence usually specifies the regulation of individual genes, but the nature of *cis*-acting information that controls genes over large regions has been elusive: higher-order DNA structure, specific histone modifications, subnuclear compartmentalization and primary DNA sequence are possibilities. One paradigm of chromosome-wide gene regulation is *Caenorhabditis elegans* dosage compensation in which a large dosage compensation complex (DCC) is targeted to both X chromosomes of hermaphrodites to repress transcript levels by half⁶. This essential process equalizes X-linked gene expression between the sexes (XO males and XX hermaphrodites). Here we report the discovery and dissection of *cis*-acting sites that mark nematode X chromosomes as targets for gene repression by the DCC. These *rex* (recruitment element on X) sites are widely dispersed along X and reside in promoters, exons and intergenic regions. *rex* sites share at least two distinct motifs that act in combination to recruit the DCC. Mutating these motifs severely reduces or abolishes DCC binding *in vivo*, demonstrating the importance of primary DNA sequence in chromosome-wide regulation. Unexpectedly, the motifs are not enriched on X, but altering motif numbers within *rex* sites demonstrates that motif co-occurrence in unusually high densities is essential for optimal DCC recruitment. Thus, X-specific repression is established through sequences not specific to X. The distribution of common motifs provides the foundation for repression along an entire chromosome.

The *C. elegans* DCC resembles condensin, a conserved protein complex required for mitotic and meiotic chromosome compaction, resolution and segregation^{7–9}. Participation of condensin and condensin-like proteins in gene regulation extends beyond dosage compensation, to transcriptional silencing in yeast¹⁰ and position-effect variegation in flies^{11,12}, suggesting a general role for these proteins in establishing and maintaining a repressed state.

Our goals were to identify X chromosome sites responsible for DCC recruitment and to define molecular features within these *rex* sites critical for target recognition and DCC binding. Previous work surveyed large regions of X (1–10 million base pairs (Mbp)) for their abilities to recruit the DCC when detached from X¹³. Some detached regions recruited the DCC but others did not, yet the DCC localized to all corresponding regions of the intact X (Fig. 1a), suggesting that recruitment sites are widely distributed along X to bind the DCC and nucleate DCC spreading to X regions lacking recruitment sites¹³.

To identify specific *rex* sites, we generated transgenic nematode lines carrying extrachromosomal arrays containing multiple copies

of individual X-chromosome cosmids sampled from large DCC-recruiting regions (Fig. 1a). DCC recruitment to arrays was examined in intestinal cell nuclei by using fluorescence *in situ* hybridization to mark array position and antibodies against dosage compensation proteins to mark DCC location¹³ (Supplementary Fig. 1a). The first four positive cosmids identified were dissected further by assaying successively smaller sub-cosmid fragments to define the minimal DNA fragments, and ultimately the shared *cis*-acting motifs, responsible for optimal DCC binding.

Full recruitment ability was ascribed to 241 base pairs (bp) for *rex-1*, 147 bp for *rex-2*, 115 bp for *rex-3*, and 411 bp for *rex-4* (Table 1, Fig. 1a, b, and Supplementary Methods). Arrays containing only 33 bp of *rex-1* recruited the DCC robustly in all nuclei, although arrays of larger *rex-1* fragments (60–241 bp) had stronger DCC recruitment activity (Fig. 2). No bias was found in the location of *rex* sites relative to coding sequences: both *rex-1* and *rex-2* reside in intergenic regions, *rex-3* in an exon, and *rex-4* in a promoter. Furthermore, *rex* sites seem widely spaced: *rex-3* is flanked by at least 300 kb of non-recruiting DNA and *rex-4* by a minimum of 100 kb.

DCC binding to *rex* arrays mimics DCC binding to intact X chromosomes. First, binding to *rex-1* requires SDC-2, the hermaphrodite-specific dosage compensation protein that triggers the binding of all DCC proteins to X and binds X without other DCC components, notably SDC-3, its partner in recruitment^{6,14}. In *sdc-2* mutants, neither DCC subunit SDC-3 nor DPY-27 was detected on arrays (Supplementary Fig. 1a, b). Furthermore, in *sdc-3* mutants, SDC-2—but not DPY-27—bound *rex-1* arrays (Supplementary Fig. 1c). Second, *rex-1* sites integrated at low copy number into autosomes recruit the DCC (Supplementary Fig. 1d). No significant DCC spreading was detected in flanking autosomal chromatin, perhaps indicating the absence of DNA sequences required for spreading or, alternatively, the presence of nearby inhibitory DNA sequences, proteins or chromatin modifications.

Sequence analysis of *rex-1* to *rex-4* with the use of MEME¹⁵ revealed multiple occurrences of two short degenerate DNA motifs clustered within each *rex* site, namely motif A (C/GCAGGGG) and motif B (T/GGTAATTG) (Fig. 1b, c, Methods, Supplementary Table 1 and Supplementary Methods). A and B motifs occur in the four *rex* sites (totalling 914 bp) at frequencies more than 20-fold greater ($P < 10^{-10}$) than empirically determined whole-genome frequencies. In contrast, the motifs are not significantly overrepresented in 16 X-linked cosmids that failed to recruit the DCC (totalling 535 kb) or in a set of smaller non-recruiting X fragments (totalling 23 kb) similar in size to *rex* sites. Despite occurrences at unusually high frequencies in *rex* sites, A motifs are not found more frequently on X than on autosomes, and B motifs are only slightly (1.25-fold) enriched on X.

¹Howard Hughes Medical Institute and ²Department of Molecular and Cell Biology, University of California – Berkeley, 16 Barker Hall Berkeley, California 94720-3204, USA.

*These authors contributed equally to this work.

Computational comparisons showed the overall density of 600-bp windows containing both A and B motifs is a strong discriminator ($P < 0.003$) between *rex* cosmids (higher density) and non-recruiting X cosmids or autosomes or even the entire X chromosome (all lower density). The discrimination between these data sets is even more apparent ($P < 0.001$) when the windows are assessed by the sum of scores for the two motifs (Supplementary Methods). Among many models tested, one that best discriminates between recruiting and non-recruiting cosmids in our training set (4 *rex* cosmids and 16 non-*rex* cosmids) predicts that a cosmid with at least one A and one B motif, each with a raw motif score of 8.0 or more and clustered within a 600-bp window, should recruit the DCC.

Applying these parameters to 30-kb segments tiling the entire X, we observed a non-uniform distribution of predicted positive segments among known¹³ strongly recruiting, weakly recruiting and non-recruiting regions of X (Fig. 1d, and Supplementary Fig. 2). Recruiting regions are significantly enriched for predicted positive windows relative to non-recruiting regions ($P < 0.004$; Supplementary Table 2). Thus, the distribution of high-scoring motif A and B clusters on X corresponds well to previous¹³ DCC recruitment studies.

The utility of this model was assessed by predicting and assaying recruitment abilities of individual cosmids covering a 2-Mbp region of X, including region D and part of C (Fig. 1a) but excluding the training set. Four of nine positive cosmid predictions proved correct, demonstrating a true positive prediction frequency of 0.44. In addition, 27 of 34 predictions correctly identified non-recruiting cosmids. The probability of selecting a true positive cosmid in this region by chance is 0.22, because 13 of 58 total cosmids (including

the training set) recruited. Thus, even this simple model using high-scoring A and B motif clusters improves positive cosmid prediction.

The functional significance of A and B motifs in DCC recruitment was established through mutational analysis of *rex* sites (Table 1, Fig. 2, Supplementary Table 1 and Supplementary Figs 3 and 4). (See Table 1 legend for nomenclature.) Recruitment *in vivo* was judged by two parameters: percentage of array-bearing nuclei with a DCC-recruiting array, and strength of DCC recruitment to arrays compared with recruitment to X chromosomes. Five recruitment strengths could be distinguished qualitatively (Fig. 2). Category 0 arrays never recruited the DCC. Category 1 arrays had infrequent, patchy DCC co-localization; X staining was bright. Category 2 arrays had consistently robust DCC staining but not significantly brighter than X staining. Category 3 arrays recruited so strongly they competed with X for the limited pool of DCC and appeared brightly stained relative to the faint (category 3⁻) or undetectable (category 3) X chromosomes. Wild-type *rex-1*•241, *rex-3*•115 and *rex-4*•411 arrays are of category 3 (100% recruiting nuclei); wild-type *rex-1*•33 and *rex-2*•147 arrays are of category 2 (100% recruiting nuclei).

Motif A is critical for DCC recruitment (Table 1). Mutating six nucleotides of the single A motif in either the 33-bp *rex-1* fragment (*rex-1*•33mA₁), the 147-bp *rex-2* fragment (*rex-2*•147mA₁) or the 115-bp *rex-3* fragment (*rex-3*•115mA₁) abolished DCC recruitment (category 0; Fig. 2b, and Supplementary Figs 3 and 4b, c). In fact, DCC recruitment to *rex-1* was severely impaired by just a 2-bp substitution in motif A (category 1, 17% recruiting nuclei; Supplementary Fig. 4b). *rex-4* has two A motifs, and recruitment was disrupted only when both motifs (*rex-4*•411mA₁A₂) were mutated (category 2, 20% recruiting nuclei; Supplementary Fig. 4c).

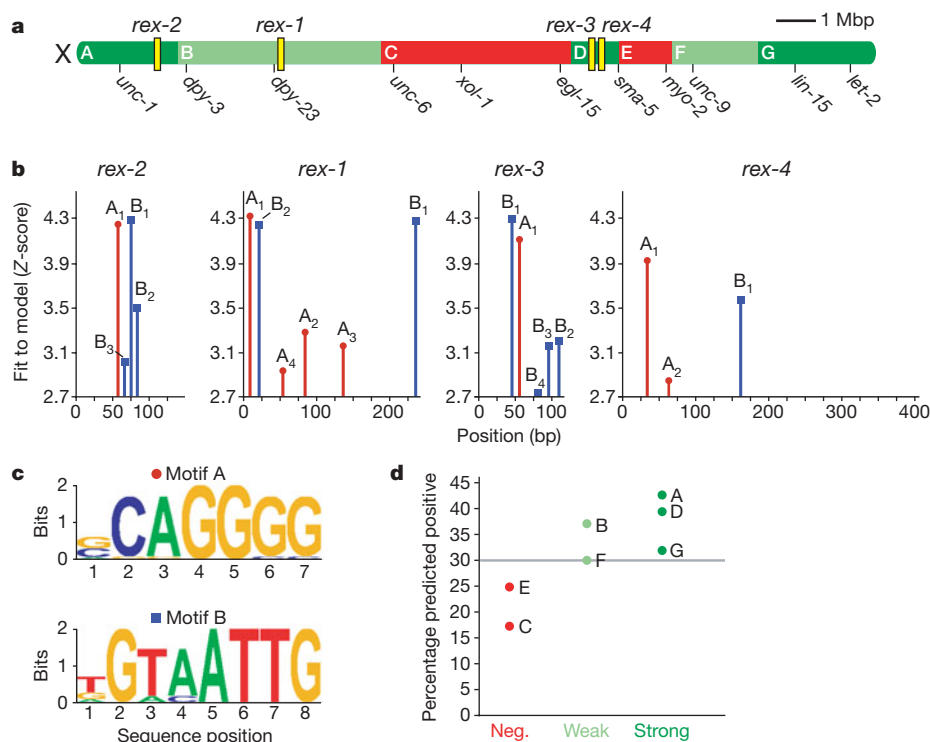


Figure 1 | DCC recruitment elements on X (*rex* sites) contain clusters of *cis*-acting regulatory motifs. **a**, DCC recruitment map of the *C. elegans* X chromosome. The positions of dissected *rex* sites (yellow) are indicated relative to previously determined X regions¹³ that strongly recruit (dark green), weakly recruit (light green) or fail to recruit (red) the DCC when detached from X. **b**, Positions of *cis*-regulatory motifs A (red circles) and B (blue squares) and their corresponding Z-scores in each full-strength *rex* site (241 bp for *rex-1*, 147 bp for *rex-2*, 115 bp for *rex-3*, and 411 bp for *rex-4*). Z-scores for every nucleotide position were calculated as (raw score minus mean score)/(standard deviation of all scores). Raw scores were generated by comparing every window of seven or eight nucleotides to position weight

matrices (PWMs) for motif A and motif B. Within a *rex* site, individual A and B motifs were named according to their relative scores against the PWMs, with A₁ and B₁ being the best scoring instances per site (see Supplementary Table 1). **c**, Models describing *cis*-acting regulatory motifs A and B based on PWMs. **d**, Plot showing correspondence between a predictive model for DCC recruitment (see the text) and previous¹³ X-chromosome-wide recruitment data. The percentage of predicted positive 30-kb segments is significantly greater in strongly (A, D and G; dark green) and weakly (B and F; light green) recruiting X regions than in non-recruiting regions (C and E; red) ($P < 0.004$; see Supplementary Table 2). The percentage of positive windows predicted for the entire X is 30% (grey line).

Motif B is also important for DCC recruitment (Table 1). Altering the single B motif in *rex-1* (*rex-1*•33mB₂) nearly eliminated DCC binding (category 1, 9% recruiting nuclei), and altering the single B motif in *rex-4* (*rex-4*•41mB₁) severely reduced binding (category 2, 32% recruiting nuclei; Supplementary Fig. 4d). In contrast, mutating six nucleotides not overlapping either motif A or B in *rex-1*•33 (*rex-1*•33m control) had little effect (category 2, 96% recruiting nuclei), confirming specificity in disrupting DCC binding (Supplementary Fig. 4b).

Strength of DCC recruitment to *rex-1* fragments is correlated with motif number. Arrays carrying the minimal fragment *rex-1*•33 (1A,1B motifs) exhibited category 2 recruitment with obvious X staining, indicating the inability to outcompete X for DCC binding, whereas arrays of *rex-1*•241 (4A,2B motifs) exhibited category 3 recruitment with undetectable X staining, indicating that X was deprived of DCC binding (Table 1 and Fig. 2a, f). This distinction was evident in both polyploid cells of adult intestines (Fig. 2) and

diploid cells of young embryos (Supplementary Fig. 5), when dosage compensation is first established. Thus, *rex-1*•241 arrays should disrupt dosage compensation to a greater extent than *rex-1*•33 arrays, a prediction confirmed by genetic assays below.

Mutations in the sex-determination gene *xol-1* cause complete male lethality from inappropriate DCC binding to the single male X and the consequent reduction of gene expression^{7,16}. *xol-1* mutant males can be rescued by disrupting DCC components¹⁶ or by titrating the DCC from X with a new target such as a *rex* array¹³. The extent of male rescue should correlate with the target's ability to compete with X for DCC binding. We found that *rex-1*•33 arrays (category 2) failed to rescue *xol-1* male lethality, as did *rex-1*•60 arrays (category 3⁻), but *rex-1*•148 arrays (category 3) rescued 55% of *xol-1* males, and *rex-1*•241 arrays (category 3) rescued 98% (Supplementary Methods).

Whereas reduction in DCC binding to the male X increases male viability, reduction in binding to hermaphrodite X chromosomes

Table 1 | Motifs A and B act in combination and are crucial for DCC recruitment to *rex* sites

Construct*	No. of wild-type motifs	Recruitment strength†	Recruitment (%)‡	No. of nuclei (no. of lines)§
<i>rex-1</i>				
<i>rex-1</i> cosmid R160	213A,397B	2	94	35 (2)
<i>rex-1</i> •33wt	1A,1B	2	100	>50 (2)
<i>rex-1</i> •33mA ₁	0A,1B	0	0	55 (3)
<i>rex-1</i> •33mA ₁ (GttGGGG)	0A,1B	1	17	57 (2)
<i>rex-1</i> •33mA ₁ (GCAttGG)	0A,1B	1	41	58 (3)
<i>rex-1</i> •33mA ₁ (GCAGGtt)	0A,1B	1	22	52 (3)
<i>rex-1</i> •33mB ₂ (GTACCAAA)	1A,0B	1	9	39 (2)
<i>rex-1</i> •33mControl¶	1A,1B	2	96	30 (2)
<i>rex-1</i> •60wt	2A,1B	3 ⁻	100	>50 (2)
<i>rex-1</i> •60mA ₁	1A,1B	2	100	69 (3)
<i>rex-1</i> •60mA ₁ A ₄	0A,1B	1	33	52 (2)
<i>rex-1</i> •89wt	3A,1B	3	100	47 (2)
<i>rex-1</i> •89mA ₁	2A,1B	3 ⁻	100	53 (2)
<i>rex-1</i> •148wt	4A,1B	3	100	>50 (2)
<i>rex-1</i> •148mA ₁	3A,1B	3	100	44 (2)
<i>rex-1</i> •148mA ₁ A ₃	2A,1B	3 ⁻	100	37 (2)
<i>rex-1</i> •148mA ₁ A ₃ A ₄	1A,1B	2	100	60 (3)
<i>rex-1</i> •148mA ₁ A ₂ A ₃ A ₄	0A,1B	1	36	61 (3)
<i>rex-1</i> •148mB ₂ (GTACCAAA)	4A,0B	2	100	28 (2)
<i>rex-1</i> •241wt	4A,2B	3	100	>50 (2)
<i>rex-1</i> •241mA ₁	3A,2B	3	100	43 (2)
<i>rex-1</i> •241mB ₂ (GTACCAAA)	4A,1B	2	100	68 (2)
<i>rex-2</i>				
<i>rex-2</i> cosmid B0294	154A,433B	2	100	>50 (2)
<i>rex-2</i> •147wt	1A,3B	2	100	63 (2)
<i>rex-2</i> •147mA ₁	0A,3B	0	0	41 (2)
<i>rex-3</i>				
<i>rex-3</i> cosmid F42E11	89A,331B	2	100	40 (2)
<i>rex-3</i> •115wt	1A,4B	3	100	>50 (2)
<i>rex-3</i> •115mA ₁	0A,4B	0	0	>50 (2)
<i>rex-3</i> •115mB ₁	1A,3B	3	100	>50 (3)
<i>rex-3</i> •115mB ₁ B ₂	1A,2B	3 ⁻	95	38 (3)
<i>rex-3</i> •115mB ₁ B ₂ B ₃	1A,1B	2	70	90 (3)
<i>rex-3</i> •115mB ₁ B ₂ B ₃ B ₄	1A,0B	2	15	74 (3)
<i>rex-4</i>				
<i>rex-4</i> cosmid F29G6	99A,349B	2	94	35 (2)
<i>rex-4</i> •411wt	2A,1B	3 ⁻	100	35 (2)
<i>rex-4</i> •411mA ₁	1A,1B	3 ⁻	100	>100 (2)
<i>rex-4</i> •411mA ₁ A ₂	0A,1B	2	20	24 (2)
<i>rex-4</i> •411mB ₁	2A,0B	2	32	60 (3)

The table shows the specific effects on DCC recruitment of multiple A and B motifs within each *rex* site and the effects of subsequent mutation of these motifs. *rex* arrays were scored for DCC binding by using two parameters, namely recruitment strength to an array and percentage of array-bearing nuclei with a recruiting array (described in the text). Increasing the number of motifs by a stepwise extension of *rex-1* from 33 bp (*rex-1*•33wt) through 60, 89 and 148 bp to 241 bp (*rex-1*•241wt) improved both recruitment strength and percentage recruitment. The enhanced recruitment was abolished on subsequent mutation of the added motifs, revealing the central role of these motifs in DCC recruitment. These results confirm the functional importance of A and B motifs in DCC recruitment and highlight the role of motif density. The *rex-1* fragments with identical numbers of A and B motifs exhibited similar recruitment strengths. Arrays from *rex-1* fragments with one A motif and one B motif (*rex-1*•33, *rex-1*•60mA₁ and *rex-1*•148mA₁A₃A₄) were in category 2; those with two A motifs and one B motif (*rex-1*•60, *rex-1*•89mA₁ and *rex-1*•148mA₁A₃) were in category 3⁻ with faint X staining; and those with three A motifs and one B motif (*rex-1*•148mA₁ and *rex-1*•89) were in category 3 with undetectable X staining.

* Nomenclature for *rex* sites is exemplified by *rex-1*•148mA₁A₃A₄, which indicates that motifs A₁, A₃ and A₄ have been mutated in the 148-bp *rex-1* fragment. 'm' precedes the complete list of mutant motifs. For mutational analysis, positions 2–7 of motif A (G/CCAGGG) were changed to TTTTTT, and positions 2–8 of motif B (T/GTAATTG) were changed to TTTTTT, except where specifically noted in parentheses. Sequences of all wild-type and mutant motifs are given in Supplementary Table 1.

† Strength of DCC recruitment to an array compared with recruitment to X chromosomes. Category 0, never recruit DCC to array; category 1, infrequent and patchy DCC localization to array; category 2, robust DCC recruitment to all arrays with staining not brighter than X staining; category 3⁻, robust DCC recruitment to all arrays with faint X staining; category 3, robust DCC recruitment to all arrays with undetectable X staining.

‡ Percentage of array-bearing nuclei with a DCC recruiting array.

§ Total number of array-bearing nuclei scored. The number of independent transgenic lines is given in parenthesis.

|| No recruitment in one line and very weak recruitment in the other.

¶ In this control, GCTGCG was changed to TTATTT.

should lower hermaphrodite viability. Indeed, 31% of hermaphrodites carrying *rex-1*•241 arrays died from defective dosage compensation, and escapers exhibited a partial disruption of dosage compensation. In contrast, only 10% of *rex-1*•33-bearing hermaphrodites died, and escapers seemed normal. These combined genetic results corroborate our cytological evidence that *rex-1* fragments with more motifs elicit stronger DCC recruitment.

To determine whether the enhanced recruitment strength of *rex-1*•241 compared with that of *rex-1*•33 was due to additional A motifs, we assayed a series of *rex-1* fragments of increasing size, such that

each progressively larger fragment included one additional A motif. Contributions of specific A motifs were then verified by mutational analysis (Table 1, Fig. 2, Supplementary Table 1 and Supplementary Fig. 4a, b). Extending *rex-1* from 33 bp to 60 bp (*rex-1*•60, with 2A,1B motifs) added one A motif (A_4) and markedly increased DCC recruitment strength from category 2 to 3⁺. Mutating an A motif (A_1) in *rex-1*•60 to generate *rex-1*•60m A_1 (1A,1B motifs) returned recruitment strength to category 2, the strength of *rex-1*•33. Mutating the remaining A motif (A_4) to generate *rex-1*•60m A_1A_4 (0A,1B motifs) reduced recruitment strength further, to category 1, establishing that A_4 was responsible for the enhanced DCC binding to *rex-1*•60 compared with *rex-1*•33.

Increasing the number of A motifs in *rex-1* to three (*rex-1*•89, with 3A,1B motifs) or four (*rex-1*•148, with 4A,1B motifs) strengthened recruitment to category 3, making it indistinguishable from that of *rex-1*•241 (4A,2B motifs). As expected, mutating one motif (A_1) in *rex-1*•89 to generate *rex-1*•89m A_1 (2A,1B motifs) and mutating two motifs (A_1 and A_3) in *rex-1*•148 to yield *rex-1*•148m A_1A_3 (2A,1B motifs) reduced recruitment strength to that of *rex-1*•60 (2A,1B motifs; category 3⁺). Mutating a third motif (A_4) in *rex-1*•148 (*rex-1*•148m $A_1A_3A_4$, with 1A,1B motifs) reduced recruitment strength to category 2 (100% recruiting nuclei), and mutating the fourth motif (A_2) to generate *rex-1*•148m $A_1A_2A_3A_4$ (0A,1B motifs) reduced recruitment strength to category 1 (36% recruiting nuclei).

In summary, *rex-1* fragments with identical numbers of A and B motifs had similar recruitment strengths, and DCC recruitment to *rex-1* was markedly improved by the increased occurrence of A motifs. The degree of improvement in DCC binding indicates that A motifs might act cooperatively.

B motifs seem additive in function. Disrupting the highest-scoring B motif in *rex-3* (*rex-3*•115m B_1 , with 1A,3B motifs) had no discernible effect on DCC binding, disrupting two B motifs (*rex-3*•115m B_1B_2 , with 1A,2B motifs) had a minor effect, disrupting three B motifs (*rex-3*•115m $B_1B_2B_3$, with 1A,1B motifs) caused substantial reduction, and disrupting all four B motifs eliminated most DCC binding (Table 1 and Supplementary Fig. 4d). These results, together with the enhanced recruitment to *rex-1* conferred by multiple A motifs, show that although a single A and B motif pair is sufficient to attract the DCC in our array assay (as in *rex-1*•33), optimal recruitment is attained only when multiple motifs are present (as in *rex-1*•241 and *rex-3*•115). On the native X chromosome, a recruitment site with a higher density of motifs, even those with lower scores, is more likely to attract the DCC than a site with a lower motif density.

Although A and B motifs act in combination, A motifs can compensate, although not completely, for the loss of motif B (Table 1 and

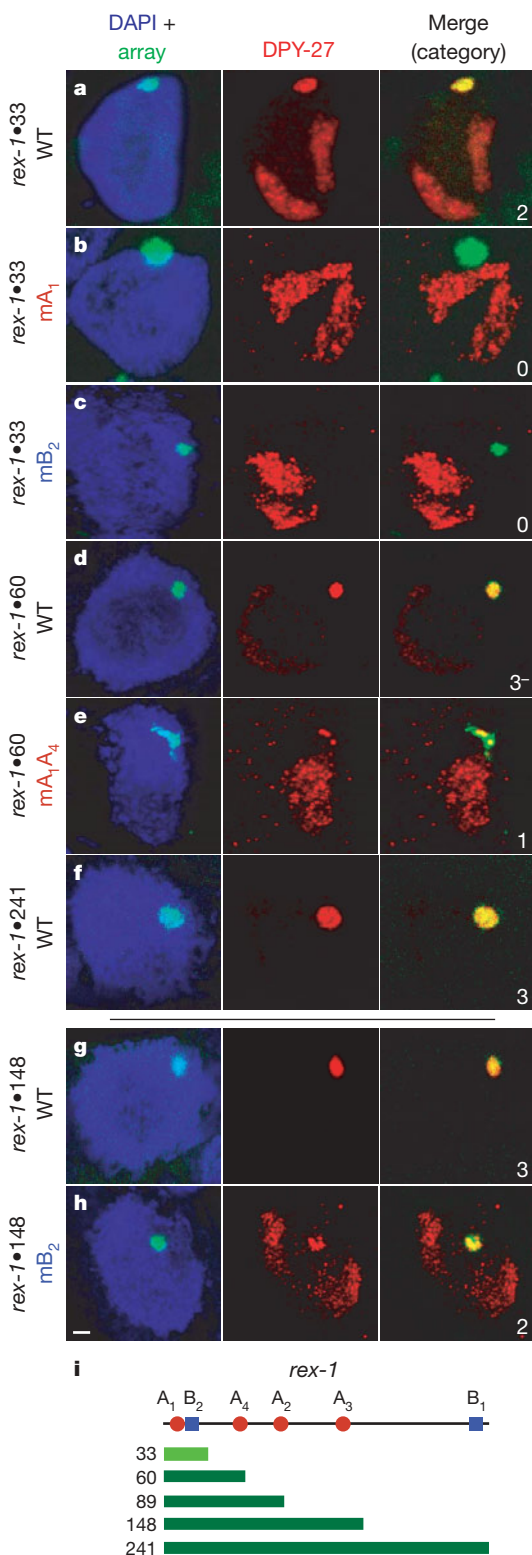


Figure 2 | Mutational analysis of *rex-1* establishes motifs A and B as cis-acting regulatory elements critical for DCC recruitment. Confocal images of intestinal cell nuclei (4,6-diamidino-2-phenylindole (DAPI) stain, blue) carrying wild-type (WT; **a**, **d**, **f**, **g**) or mutant (**b**, **c**, **e**, **h**) *rex* arrays (fluorescence *in situ* hybridization, green) co-stained with DPY-27 antibodies (red). DPY-27 binding to X requires all other known DCC components. The recruitment strength category (see the text for description) of each array is indicated by a number (0–3) at the right.

a, Arrays containing a cloned 33-bp *rex-1* fragment (1A,1B motifs) recruit the DCC robustly. **b**, **c**, Mutating either motif A (**b**) or motif B (**c**) in *rex-1*•33 abolishes DCC recruitment. (See also Supplementary Fig. 1.) **d**, Extending *rex-1* to 60 bp adds a second motif A and markedly increases recruitment strength to the degree that arrays begin to outcompete the X for DCC binding, as demonstrated by the weaker X chromosome staining relative to array staining. **e**, When both A motifs in *rex-1*•60 are mutated, DCC recruitment to the arrays is severely reduced, and X chromosome staining is restored. **f**, **g**, *rex-1*•241 (**f**) and *rex-1*•148 (**g**) each contain four A motifs as well as two or one B motifs, respectively; both exhibit maximum recruitment strength, completely outcompeting the X for DCC staining. **h**, Loss of the only B motif in *rex-1*•148 (*rex-1*•148 m B_2) reduces recruitment but does not eliminate it. **i**, Map of *rex-1* fragments used in this paper. Red circles represent A motifs; blue squares represent B motifs. Stronger recruitment is indicated by darker green. Scale bar, 2 μ m.

Figs 2c, g, h). Mutation of the single B in *rex-1*•33 (*rex-1*•33mB₂, with 1A,0B motifs) nearly eliminated DCC binding (category 1, 9% recruitment). However, the addition of three A motifs, as in *rex-1*•148mB₂ (4A,0B motifs), partly restored binding (category 2, 100% recruitment). The B₂ mutation could not be suppressed further by extending *rex-1* to include an additional B motif (*rex-1*•241mB₂, with 4A,1B motifs), indicating at least some dependence of DCC recruitment on relative motif positions as well as absolute motif number.

Our cumulative results show that at least two disparate DNA motifs, probably recognized by different proteins (or possibly RNAs) act in combination to recruit the DCC. The involvement of co-occurring motifs increases the specificity and potentially the stability of DCC binding. Furthermore, each motif class provides a separate opportunity to enlist regulatory factors in X repression. Despite the participation of two motifs in recruitment, multiple A motifs can recruit the DCC in part even without motif B. Cooperativity in DCC binding to neighbouring A motifs could account for the potency of multiple A motifs in DCC recruitment and their partial ability to overcome the loss of motif B. Alternatively, if a motif-B-binding factor is essential for DCC recruitment, it might be brought to *rex* sites through association with motif-A-binding factors and this interaction stabilized through motif B binding.

Our work has decoded essential *cis*-acting information that directs the *C. elegans* dosage compensation complex to repress X-chromosome gene expression. Remarkably, X-specific repression is established through sequences not specific to the X chromosome. *rex* sites enriched for the co-occurrence of at least two prevalent motifs mark X chromosomes as targets for DCC binding. Additional features, possibly including chromatin modifications or other regulatory sequences, might synergize with *rex* sites to permit DCC binding on X and prevent it on autosomes. Widely dispersed *rex* sites, in the context of X, recruit the DCC and probably nucleate DCC spreading to neighbouring areas that cannot independently bind the DCC. In this manner, the arrangement of common motifs induces chromosome-wide repression.

Recent studies of dosage compensation in *Drosophila melanogaster* identified X fragments bound by the MSL (male-specific lethal) complex^{17–20}, a male-specific RNA–protein complex that increases transcription from the single male X chromosome. Computational analysis^{17,18} of these fragments, bound through initial MSL recruitment or subsequent MSL spreading, identified many combinations of shared, degenerate motifs, but none were specific to X. If mutational analysis proves the fly motifs to be essential for MSL binding, as we have shown for worm motifs in DCC recruitment, remarkably similar principles would seem to govern the X-specific binding of two evolutionarily unrelated protein complexes that regulate whole chromosomes in opposite ways.

METHODS

Motifs. Motif finding was performed with MEME¹⁵ on the four *rex* sequences. Starting parameters were selected to maximize the recovery of motifs that conform to a priori expectations for binding sites of sequence-specific DNA-binding proteins, including degenerate DNA motifs 6–16 bp in length, occurring on either strand, often repeated multiple times in a local region of DNA. From models trained under these initial conditions, the two most consistently high-scoring candidate motifs, A and B, were chosen for functional analysis (Supplementary Methods). Patser²¹ was used to scan all sequence data sets against position weight matrices for motifs A and B.

Recruitment sites. *rex* sequences reside at the following X chromosome coordinates: *rex-1*•241, X:4395434,4395674; *rex-2*•147, X:1908940,1909087; *rex-3*•115, X:11361205,11361319; and *rex-4*•411, X:11521745,11522155. They can be downloaded from <http://www.wormbase.org>. Nucleotide sequences of smaller *rex-1* fragments are given in Supplementary Methods.

Received 10 July; accepted 11 October 2006.

Published online 19 November 2006.

- Grewal, S. I. S. & Elgin, S. C. R. Heterochromatin: new possibilities for the inheritance of structure. *Curr. Opin. Genet. Dev.* **12**, 178–187 (2002).
- Yamada, T., Fischle, W., Sugiyama, T., Allis, C. D. & Grewal, S. I. S. The nucleation of heterochromatin by a histone deacetylase in fission yeast. *Mol. Cell* **20**, 173–185 (2005).
- Reik, W. *et al.* Chromosome loops, insulators, and histone methylation: new insights into regulation of imprinting in clusters. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 29–37 (2004).
- Nusinow, D. A. & Panning, B. Recognition and modification of sex chromosomes. *Curr. Opin. Genet. Dev.* **15**, 91–122 (2005).
- Lucchesi, J. C., Kelly, W. G. & Panning, B. Chromatin remodeling in dosage compensation. *Annu. Rev. Genet.* **39**, 615–651 (2005).
- Meyer, B. J. X-chromosome dosage compensation. (http://www.wormbook.org/chapters/www_dosagecomp/dosagecomp.html) (2005).
- Chuang, P.-T., Albertson, D. G. & Meyer, B. J. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).
- Lieb, J. D., Albrecht, M. R., Chuang, P.-T. & Meyer, B. J. MIX-1: An essential component of the *C. elegans* mitotic machinery executes X chromosome dosage compensation. *Cell* **92**, 265–277 (1998).
- Losada, A. & Hirano, T. Dynamic molecular linkers of the genome: the first decade of the SMC proteins. *Genes Dev.* **19**, 1269–1287 (2005).
- Bhalla, N., Biggins, S. & Murray, A. W. Mutation of YCS4, a budding yeast condensin subunit, affects mitotic and nonmitotic chromosome behavior. *Mol. Biol. Cell* **13**, 632–645 (2002).
- Dej, K. J., Ahn, C. & Orr-Weaver, T. L. Mutations in the *Drosophila* condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. *Genetics* **168**, 895–906 (2004).
- Cobbe, N., Savvidou, E. & Heck, M. M. Diverse mitotic and interphase functions of condensins in *Drosophila*. *Genetics* **172**, 991–1008 (2006).
- Csankovszki, G., McDonel, P. & Meyer, B. J. Recruitment and spreading of the *C. elegans* dosage compensation complex along X chromosomes. *Science* **303**, 1182–1185 (2004).
- Dawes, H. E. *et al.* Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. *Science* **284**, 1800–1804 (1999).
- Bailey, T. L. & Elkan, C. in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (eds Altman, R., Brutlag, D., Karp, P., Lathrop, R. & Searls, D.) 28–36 (AAAI Press, Menlo Park, California, 1994).
- Miller, L. M., Plenefisch, J. D., Casson, L. P. & Meyer, B. J. *xol-1*: a gene that controls the male modes of both sex determination and X chromosome dosage compensation in *C. elegans*. *Cell* **55**, 167–183 (1988).
- Dahlsveen, I. K., Gilfillan, G. D., Shelest, V. I., Lamm, R. & Becker, P. B. Targeting determinants of dosage compensation in *Drosophila*. *PLoS Genet.* **2**, e5 (2006).
- Gilfillan, G. D. *et al.* Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev.* **20**, 858–870 (2006).
- Alekseyenko, A. A., Larschan, E., Lai, W. R., Park, P. J. & Kuroda, M. I. High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev.* **20**, 848–857 (2006).
- Legube, G., McWeeney, S. K., Lercher, M. J. & Akhtar, A. X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in *Drosophila*. *Genes Dev.* **20**, 871–883 (2006).
- Hertz, G. Z. & Stormo, G. D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563–577 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Coulson for cosmid; S. Uzawa for technical innovation with image analysis; P. Nix for help with *xol-1* XO rescue experiments; M. Ralston for statistical analysis; and R. Auty, M. Botchan, S. Brenner, T. Cline, R. Losick, E. Ralston and A. Severson for comments on the manuscript. This work was supported by an NIH grant (B.J.M.), NIH training grants (P.M. and B.K.P.), and a Netherlands Organisation for Scientific Research (NWO) Talent fellowship (J.J.). B.J.M. is an investigator and J.J. an associate of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.J.M. (bjmeyer@berkeley.edu).

Centriole assembly in *Caenorhabditis elegans*

Laurence Pelletier¹, Eileen O'Toole², Anne Schwager¹, Anthony A. Hyman¹ & Thomas Müller-Reichert¹

Centrioles are necessary for flagella and cilia formation^{1,2}, cytokinesis^{3,4}, cell-cycle control⁵ and centrosome organization/spindle assembly⁶. They duplicate once per cell cycle, but the mechanisms underlying their duplication remain unclear. Here we show using electron tomography of staged *C. elegans* one-cell embryos that daughter centriole assembly begins with the formation and elongation of a central tube followed by the peripheral assembly of nine singlet microtubules. Tube formation and elongation is dependent on the SAS-5 and SAS-6 proteins, whereas the assembly of singlet microtubules onto the central tube depends on SAS-4. We further show that centriole assembly is triggered by an upstream signal mediated by SPD-2 and ZYG-1. These results define a structural pathway for the assembly of a daughter centriole and should have general relevance for future studies on centriole assembly in other organisms.

The structure of centrioles is conserved from ancient eukaryotes to mammals^{1,7–9}. They are barrel-shaped, 100–250 nm in diameter and 100–400 nm in length, with a ninefold symmetric array of singlet, doublet or triplet microtubules. Daughter centrioles grow orthogonally to the older (mother) centrioles, but the assembly process remains mysterious. Centrioles in *C. elegans* are 150 nm in length, 100 nm in diameter and consist of a central tube surrounded by nine singlet microtubules¹⁰. Genomic and genetic analyses have identified proteins required for centriole duplication in *C. elegans*^{11–16}. These include the ZYG-1 kinase, the SAS proteins (SAS-4, SAS-5 and SAS-6) and SPD-2, all of which localize to centrioles^{14,16–23}.

C. elegans oocytes arrest in meiotic prophase I, at which point they lack centrioles. Fertilization of oocytes by sperm triggers the completion of meiosis and contributes a centriole pair (Supplementary Fig. 1). After meiosis, the embryo enters the first mitotic division, which is characterized by a series of easily distinguishable events. These include pronuclear appearance (PNA), pronuclear migration (PNM), pronuclear rotation (PNR) and metaphase. Using chemical fixation and serial section electron microscopy we were able to detect daughter centriole assembly intermediates at, but not before, PNM (Supplementary Fig. 1).

To determine whether the onset of daughter centriole assembly correlates with the recruitment of known centriole proteins, we performed mating-based assays^{18,20,22–24} (Supplementary Fig. 2). In these assays, the sperm donates an unlabelled centriole pair, whereas the green fluorescent protein (GFP)-tagged centriole proteins are contributed by the maternal cytoplasm. Their incorporation into nascent daughter centrioles can therefore be monitored (Supplementary Fig. 2). We fixed embryos at different stages of the first cell division and stained them using SAS-4 antibodies—to mark the position of the sperm centrioles—and GFP or ZYG-1 antibodies to monitor the recruitment of individual centriole proteins to the site of assembly (Fig. 1; Supplementary Fig. 3 for GFP–ZYG-1 recruitment; $n > 10$ for each protein). We observed that SPD-2 and ZYG-1 are recruited to centrioles soon after fertilization, during the completion of female meiosis (Fig. 1a), whereas SAS-4, SAS-5 and SAS-6 are recruited shortly afterwards, at PNA (Fig. 1b).

To gain insights into the assembly relationship between these proteins, we depleted individual centriole proteins by RNA-mediated interference (RNAi) and determined the effect this had on the recruitment of the other proteins. In one-cell *spd-2(RNAi)* embryos ($n = 18$), we were unable to detect ZYG-1 recruitment at any stage (Fig. 2a, top panel); however, in *zyg-1(RNAi)* embryos, SPD-2 recruitment was unaffected ($n = 14$; Fig. 2a, bottom panel). These results suggest that SPD-2 acts upstream of ZYG-1. In all *spd-2(RNAi)* embryos analysed, we were unable to detect significant recruitment of maternal SAS-4, SAS-5 or SAS-6 to the site of centriole assembly (Fig. 2b, $n > 10$ for each protein). We observed essentially the same result in *zyg-1(RNAi)* embryos (data not shown), which is in agreement with previous results^{22–24} and with our observation that SPD-2 is required to recruit ZYG-1 (Fig. 2a). In *sas-4(RNAi)* embryos, both SAS-5 and SAS-6 are recruited to the site of daughter centriole assembly, confirming published results in two-cell embryos (Fig. 2c, $n = 4$)^{22,24}. It was previously shown that SAS-4 recruitment is blocked in *sas-5(RNAi)* and *sas-6(RNAi)* embryos^{22–24}. Together, these results suggest that in one-cell embryos SPD-2 and ZYG-1 are required for daughter centriole assembly by promoting the recruitment of the centriole components SAS-5 and SAS-6 (ref. 22). SAS-5 and SAS-6 recruitment then promotes the recruitment of SAS-4.

Our thin section electron microscopy results showed that daughter centriole assembly begins during PNM (Supplementary Fig. 1). In contrast, our mating assays showed that centriole proteins are recruited beforehand, during meiosis or at PNA (Fig. 1). Two possibilities could explain these observations. First, centriole proteins may be recruited to the proximity of the mother centrioles before daughter centriole assembly. Second, the recruitment observed may coincide with the emergence of a structural intermediate that is undetectable using chemical fixation and thin section electron microscopy (Supplementary Fig. 1). To examine potential structural intermediates of centriole assembly, we developed an approach to perform high-pressure freezing and electron tomography on isolated embryos at different stages of the first cell division (for measurements see Supplementary Fig. 4 and Supplementary Table 1)²⁵. To our surprise, daughter centriole assembly intermediates were observed at PNA (Fig. 3a). The smallest structure detected was a cylindrical tube approximately 60 nm in length, similar to the central tube of mother centrioles (Fig. 3a). Electron-dense appendages emanated from the singlet microtubules of mother centrioles (Fig. 3a, bottom left panel, arrowheads). During PNM, the central tubes of daughter centrioles were longer, reaching about 110 nm at the end of PNM (Fig. 3b and Supplementary Fig. 5). We observed that the central tube of daughter centrioles increased in diameter during PNM, whereas that of the mother centriole did not (Supplementary Fig. 5); this increase coincided with the appearance of an inner/outer wall structure on daughter centriole central tubes (Fig. 3c, right panel). Singlet microtubules began assembling around the central tube during PNR and by the end of PNR nine singlet microtubules were

¹Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany. ²Boulder Laboratory for 3D Electron Microscopy of Cells, Department of Molecular, Cellular and Developmental Biology, Campus Box 347, University of Colorado, Boulder, Colorado 80309, USA.

observed (Fig. 3c). We detected hook-like appendages along the length of daughter central tubes at positions where singlet microtubules had yet to assemble (Fig. 3c, arrowheads, and data not shown). Singlet microtubule assembly did not seem to occur preferentially at distal or proximal extremities, although a slight positional bias for the

distal region of the central tube was observed (Supplementary Fig. 5). In summary, our tomography results show that in one-cell embryos, daughter centriole assembly begins with the formation of the central tube, which elongates and increases in diameter before the assembly of the nine singlet microtubules.

We looked at the effect of individually depleting centriole proteins on daughter centriole assembly. In *zyg-1(RNAi)* embryos, no daughter centriole structures were detected (Fig. 4a), consistent with the idea that in *zyg-1(RNAi)* embryos, the SAS-proteins are not recruited to the site of centriole assembly (Fig. 3b)^{22–24}. Similarly, in *sas-5(RNAi)* and *sas-6(RNAi)* embryos, no daughter centriole structures were observed (Supplementary Movie 'SAS5RNAi' and Fig. 4b, respectively). Most interestingly, in *sas-4(RNAi)* embryos at PNA central tubes of daughter centrioles still formed (Fig. 4c). After PNR, daughter centriole central tubes were longer (Supplementary Movie 'SAS4RNAi' and Supplementary Table 1), suggesting that tube elongation can still occur in *sas-4(RNAi)* embryos. Daughter centriole central tubes failed to increase in width, and seemed defective in the assembly of both singlet microtubules and hook-like appendages (Supplementary Table 1 and data not shown). In older embryos, daughter centriole central tubes were often difficult to discern (data not shown). These results are consistent with our observation for wild-type embryos that the formation of a central tube represents a structural intermediate in centriole assembly, which is unstable in *sas-4(RNAi)* embryos. Taken together, our electron tomography and light microscopy data suggest that SAS-4 is required to assemble or maintain singlet microtubules onto the central tube, the assembly of which is, in turn, dependent on SAS-5 and SAS-6.

Using electron tomography we have shown that daughter centriole assembly in one-cell *C. elegans* embryos occurs in at least three distinct steps: tube formation; tube elongation; and singlet microtubule assembly (Fig. 4d). Assembly begins at the time of PNA and is completed by the end of PNR, which we estimate requires 8–10 min (Fig. 4d). The recruitment of the SAS-proteins at PNA to the site of daughter centriole assembly is coincident with the emergence of the daughter central tube. In *spd-2(RNAi)* and *zyg-1(RNAi)* embryos, the recruitment of the SAS-proteins to the site of assembly is blocked. Accordingly, we were unable to detect daughter centriole structures in *zyg-1(RNAi)* embryos, confirming previous results¹⁴. These observations, in combination with our result showing that SPD-2 is required for the recruitment of the ZYG-1 kinase to the site of daughter centriole assembly, suggest that the role of SPD-2 in centriole duplication is to recruit ZYG-1 to centrioles (Fig. 4d). ZYG-1—possibly in concert with SPD-2—can then either function as a signal required for the initiation of daughter centriole assembly or directly drive the assembly process.

The fact that daughter centriole central tubes assemble in *sas-4(RNAi)* embryos (conditions under which SAS-5 and SAS-6 recruitment is not inhibited) indicates that SAS-5 and SAS-6 could be structural components of the central tube. SAS-5 and SAS-6 are known to physically interact, indicating that they may fulfil this structural role as a heterodimer²⁴. Alternatively, it is possible that SAS-5 and SAS-6 are required to maintain the structural integrity of the central tube, which is composed of other proteins. In

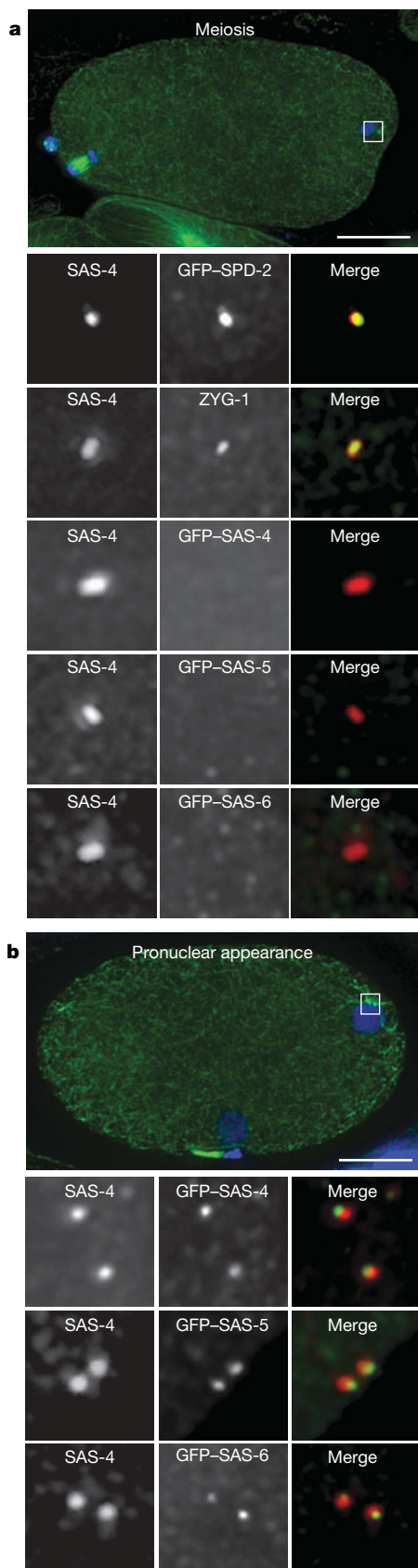


Figure 1 | Centriole proteins are recruited in two steps. **a**, Recruitment of centriole proteins during meiosis. At this stage only SPD-2 and ZYG-1 are recruited to centrioles, colocalizing with SAS-4. **b**, The SAS-proteins are recruited later, at the time of PNA. In these assays, wild-type males are mated to feminized hermaphrodites that express individual GFP-tagged centriole proteins. SAS-4 antibodies are used to label the sperm-derived centriole, therefore marking the site of daughter centriole assembly and GFP or ZYG-1 antibodies are used to monitor the recruitment of the maternal proteins. DNA (blue), microtubules (green), SAS-4 (white; red in merge) and GFP/ZYG-1 (white; green in merge) labels are shown. Given that ZYG-1 is not expressed in the sperm we used ZYG-1 antibodies to monitor the recruitment of the endogenous protein. Scale bars, 10 μ m.

sas-4(RNAi) embryos, central tube assembly is not perturbed although its stability and capacity to assemble singlet microtubules are compromised. Thus, we propose a role for SAS-4 in tethering singlet microtubules to the central tube, which would be consistent with the ring-like distribution of SAS-4 around centrioles²⁰. It was previously shown that the amount of SAS-4 on centrioles is related to the quantity of pericentriolar material they recruit²⁰. It is therefore possible that the singlet microtubules on the *C. elegans* centriole are required to recruit and organize pericentriolar material¹⁹. Recent evidence has shown that pericentriolar material is also required for daughter centriole assembly²³. It will be of interest to delineate the contribution of such proteins to daughter centriole assembly.

Published structures using chemical fixation and thin section electron microscopy on mammalian tissue culture cells are ambiguous as to whether centrioles have a central tube^{9,26,27}. It will be of importance to revisit daughter centriole assembly in mammalian tissue culture

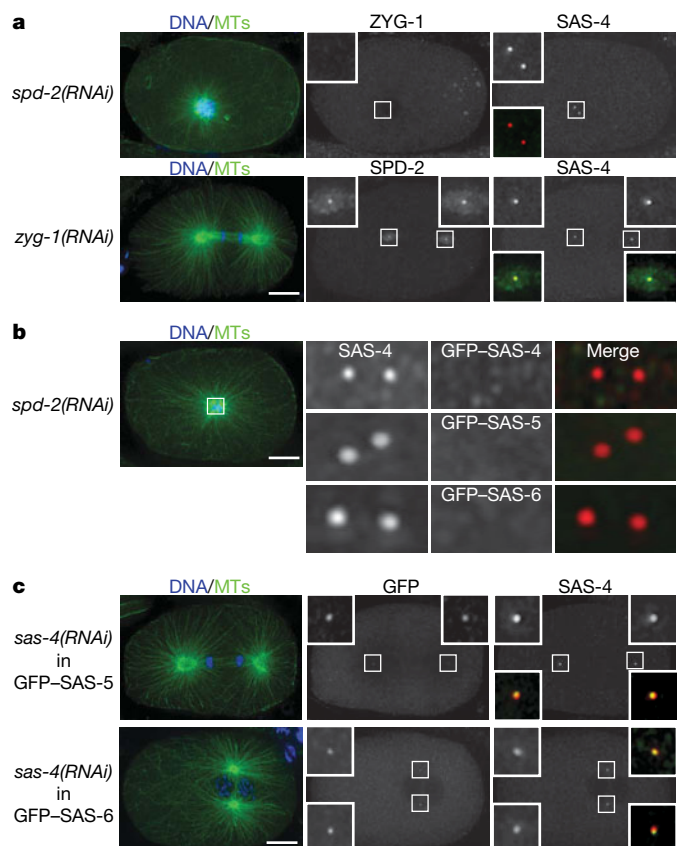


Figure 2 | SPD-2 and ZYG-1 are required for the recruitment of SAS proteins to the site of centriole assembly. **a–c**, The recruitment of individual centriole proteins was analysed in *spd-2(RNAi)*, *zyg-1(RNAi)* or *sas-4(RNAi)* embryos. **a**, In *spd-2(RNAi)* embryos ZYG-1 recruitment to centrioles is blocked (top panels). A merged image of ZYG-1 (green) with SAS-4 (red) is shown in the lower inset. In *zyg-1(RNAi)* embryos recruitment of SPD-2 is not affected (bottom panels). Merged images of SPD-2 (green) with SAS-4 (red) are shown in the lower insets. Insets are $\times 3$ magnification. **b**, In *spd-2(RNAi)* embryos, the recruitment of SAS-4, SAS-5 and SAS-6 to the site of daughter centriole assembly is blocked. Left panel shows the merged image of DNA (blue) and microtubules (green) of a representative *spd-2(RNAi)* embryo. Right panels show magnified ($\times 8$) views of SAS-4 immunolocalization; GFP-SAS-4, GFP-SAS-5 and GFP-SAS-6 labelling; and the respective merged images of regions similar to the boxed region shown in the merged image on the right. **c**, In *sas-4(RNAi)* embryos, recruitment of SAS-5 (top panel) and SAS-6 (bottom panel) still occurs. Merged images of GFP-SAS-5 (green) with SAS-4 (red) and of GFP-SAS-6 (green) with SAS-4 (red) are shown. Insets are $\times 3$ magnification. DNA (blue), microtubules (green), SAS-4 (white; red in merge) and GFP/ZYG-1/SPD-2 (white; green in merge) labels are shown. Scale bars, 10 μm .

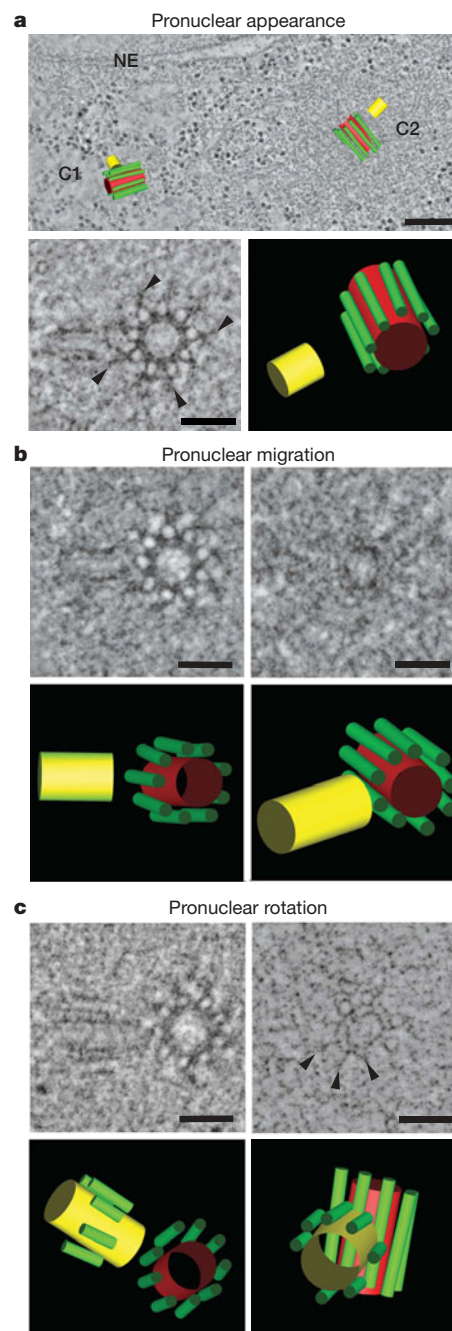


Figure 3 | Centriole assembly in *Caenorhabditis elegans* is a multi-step process. **a**, Tomographic slice and overlaid 3D model showing two centriole pairs (C1 and C2) at PNA (top). NE, nuclear envelope. Higher magnification view of another centriole pair at PNA (bottom left). The mother centriole is shown in cross-section and the assembling daughter centriole in longitudinal orientation. Note the presence of electron-dense appendages extending from the singlet microtubules of the mother centriole (arrowheads). Three-dimensional models illustrating the mother (red) and daughter (yellow) centriole central tubes and associated singlet microtubules (green) are also shown. **b**, Centrioles during PNM. The elongating central tubes of daughter centrioles are shown in both longitudinal orientation (left) and in cross section (right). **c**, Centrioles during PNR. An incomplete number of singlet microtubules are observed at this stage. Positions on the tube lacking singlet microtubules, but decorated with hook-like appendages, are indicated (arrowheads). Quicktime movies of all tomograms and 3D models can be found in Supplementary Information. Scale bars, 1 μm (top panel in **a**) and 100 nm (bottom panel in **a**, **b**, **c**).

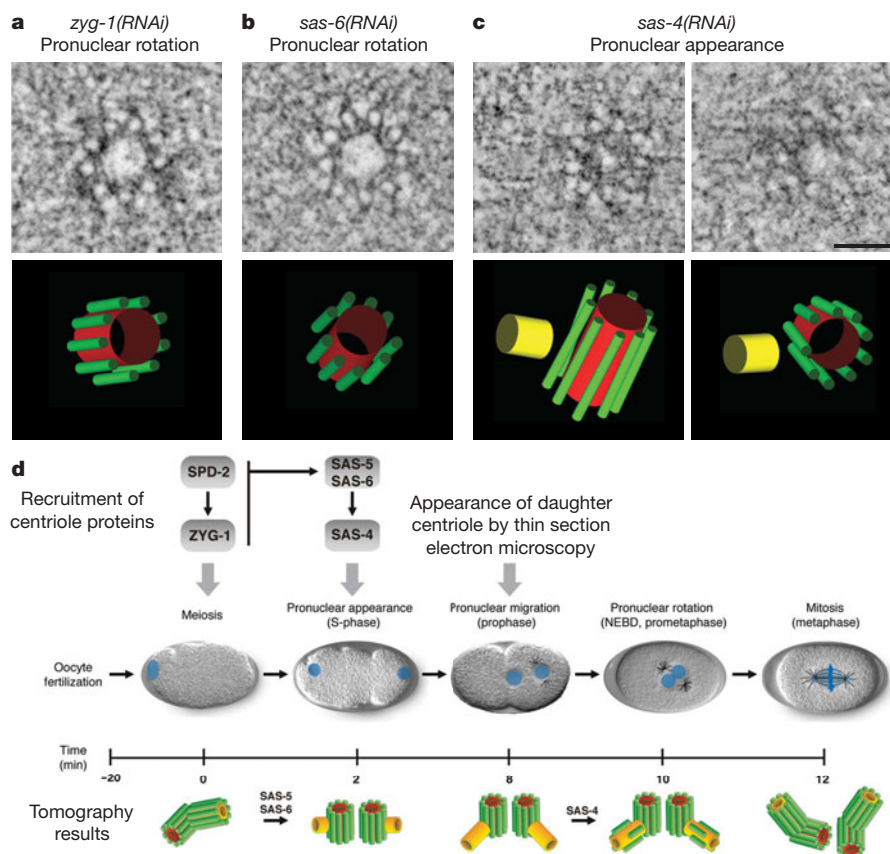


Figure 4 | The SAS proteins are required at different steps during centriole assembly. **a–c**, Tomographic reconstruction of *zyg-1(RNAi)* embryos during PNR (**a**), *sas-6(RNAi)* embryos during PNR (**b**), and *sas-4(RNAi)* embryos at PNA (**c**). Note the absence of daughter centriole structures in both *zyg-1(RNAi)* and *sas-6(RNAi)* embryos; but the presence of a daughter centriole central tube in *sas-4(RNAi)* embryos. Three-dimensional models illustrating the mother (red) and daughter (yellow) centriole central tubes and associated singlet microtubules (green) are shown in the bottom panels. Quicktime movies of all tomograms and 3D models can be found in Supplementary Information. **d**, Schematic representation of daughter centriole assembly during the first mitotic division. SPD-2 and ZYG-1 are

recruited during meiosis, before daughter centriole assembly. SPD-2 is required to recruit ZYG-1 to mother centrioles. Both SPD-2 and ZYG-1 are required for the recruitment of the SAS proteins at the time of PNA, coincident with the formation of the daughter centriole central tube. SAS-5 and SAS-6 are required for SAS-4 recruitment. In *sas-5(RNAi)* and *sas-6(RNAi)* embryos the central tubes fail to assemble. The central tube elongates during PNM and singlet microtubules assemble around it during PNR in a SAS-4 dependent fashion so that by metaphase two fully formed daughter centrioles are observed. NEBD, nuclear envelope breakdown. Scale bar, 100 nm.

cells, and other organisms, using high-pressure freezing and time-resolved tomography to determine if daughter centriole assembly also proceeds in the same order: tube formation, elongation and microtubule assembly. Interestingly, three of the five components required for centriole duplication in *C. elegans*—SPD-2, SAS-4 and SAS-6—have mammalian homologues^{17,18,23,24}, although only human SAS-6 is known to be required for centriole duplication. It is therefore likely that some of the assembly intermediates uncovered here in *C. elegans* are conserved in mammals and other eukaryotes.

Note added in proof: While this paper was being reviewed, another study²⁹ reported the sequential recruitment of *C. elegans* centriole proteins during duplication.

METHODS

***Caenorhabditis elegans* strains, RNAi and immunofluorescence.** The worm strains, antibodies, dsRNA and RNAi conditions used in this study are detailed in Supplementary Information. For immunofluorescence microscopy embryos were labelled and imaged as previously described¹⁷. Three-dimensional data sets were acquired on a DeltaVision RT imaging system (Applied Precision) and shown as maximal projections.

Specimen preparation for electron tomography. Hermaphrodites were dissected in M-9 buffer containing 20% BSA (weight by volume; Sigma), individual embryos collected into capillary tubing, and development observed using phase

contrast microscopy. Embryos at specific stages were transferred to specimen holders and frozen using the EMPACT2+RTS high-pressure freezer (Leica). Embryos were freeze-substituted, stained and imaged by electron tomography using a TECNAI F30 intermediate-voltage electron microscope (FEI) operated at 300 kV as previously described²⁵.

Modelling and analysis of tomographic data. We recorded five double-tilt data sets of poles at PNA, five at PNM, five during PNR and six during mitosis. In addition, we acquired two data sets of *zyg-1(RNAi)* embryos, four of *sas-4(RNAi)* embryos, two of *sas-5(RNAi)* embryos and three of *sas-6(RNAi)* embryos. Tomography and image analysis was carried out using the IMOD software package²⁸. Using serial slices extracted from the tomograms, we modelled the central tube and the singlet microtubules of mother and daughter centrioles. The structure of assembly intermediates was analysed by extracting a slice of image data 1-voxel thick. The orientation was adjusted to visualize the centriole in either longitudinal orientation or in cross-section. The projections of the 3D models were displayed and rotated to study their 3D geometry. To display in 3D, substructures of the centriole were meshed using the IMODMESH program and shown as tubular graphical objects. Measurements of centriole components were extracted from model contour data using the companion program, IMODINFO (Supplementary Fig. 6 and 7).

Received 25 August; accepted 6 October 2006.

1. Dutcher, S. K. Elucidation of basal body and centriole functions in *Chlamydomonas reinhardtii*. *Traffic* **4**, 443–451 (2003).

2. Briggs, L. J., Davidge, J. A., Wickstead, B., Ginger, M. L. & Gull, K. More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr. Biol.* **14**, R611–R612 (2004).
3. Gromley, A. *et al.* Centriolin anchoring of exocyst and SNARE complexes at the midbody is required for secretory-vesicle-mediated abscission. *Cell* **123**, 75–87 (2005).
4. Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291**, 1550–1553 (2001).
5. Doxsey, S., McCollum, D. & Theurkauf, W. Centrosomes in cellular regulation. *Annu. Rev. Cell Dev. Biol.* **21**, 411–434 (2005).
6. Sluder, G. & Rieder, C. L. Centriole number and the reproductive capacity of spindle poles. *J. Cell Biol.* **100**, 887–896 (1985).
7. Azimzadeh, J. & Bornens, M. in *Centrosomes in Development and Disease* (ed. Nigg, E. A.) 93–122 (Wiley-VCH, Weinheim, 2004).
8. Dippell, R. V. The development of basal bodies in paramecium. *Proc. Natl Acad. Sci. USA* **61**, 461–468 (1968).
9. Vorobjev, I. A. & Chentsov Yu, S. Centrioles in the cell cycle. I. Epithelial cells. *J. Cell Biol.* **93**, 938–949 (1982).
10. O'Toole, E. T. *et al.* Morphologically distinct microtubule ends in the mitotic centrosome of *Caenorhabditis elegans*. *J. Cell Biol.* **163**, 451–456 (2003).
11. Gönczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
12. Sönnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462–469 (2005).
13. Zipperlen, P., Fraser, A. G., Kamath, R. S., Martinez-Campos, M. & Ahringer, J. Roles for 147 embryonic lethal genes on *C. elegans* chromosome I identified by RNA interference and video microscopy. *EMBO J.* **20**, 3984–3992 (2001).
14. O'Connell, K. F. *et al.* The *C. elegans zyg-1* gene encodes a regulator of centrosome duplication with distinct maternal and paternal roles in the embryo. *Cell* **105**, 547–558 (2001).
15. O'Connell, K. F., Leys, C. M. & White, J. G. A genetic screen for temperature-sensitive cell-division mutants of *Caenorhabditis elegans*. *Genetics* **149**, 1303–1321 (1998).
16. O'Connell, K. F., Maxwell, K. N. & White, J. G. The *spd-2* gene is required for polarization of the anteroposterior axis and formation of the sperm asters in the *Caenorhabditis elegans* zygote. *Dev. Biol.* **222**, 55–70 (2000).
17. Pelletier, L. *et al.* The *Caenorhabditis elegans* centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Curr. Biol.* **14**, 863–873 (2004).
18. Leidel, S. & Gönczy, P. SAS-4 is essential for centrosome duplication in *C. elegans* and is recruited to daughter centrioles once per cell cycle. *Dev. Cell* **4**, 431–439 (2003).
19. Leidel, S. & Gönczy, P. Centrosome duplication and nematodes: recent insights from an old relationship. *Dev. Cell* **9**, 317–325 (2005).
20. Kirkham, M., Müller-Reichert, T., Oegema, K., Grill, S. & Hyman, A. A. SAS-4 is a *C. elegans* centriolar protein that controls centrosome size. *Cell* **112**, 575–587 (2003).
21. Kemp, C. A., Kopish, K. R., Zipperlen, P., Ahringer, J. & O'Connell, K. F. Centrosome maturation and duplication in *C. elegans* require the coiled-coil protein SPD-2. *Dev. Cell* **6**, 511–523 (2004).
22. Delattre, M. *et al.* Centriolar SAS-5 is required for centrosome duplication in *C. elegans*. *Nature Cell Biol.* **6**, 656–664 (2004).
23. Dammermann, A. *et al.* Centriole assembly requires both centriolar and pericentriolar material proteins. *Dev. Cell* **7**, 815–829 (2004).
24. Leidel, S., Delattre, M., Cerutti, L., Baumer, K. & Gönczy, P. SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nature Cell Biol.* **7**, 115–125 (2005).
25. Müller-Reichert, T., Srayko, M., Hyman, A., O'Toole, E. & McDonald, K. Correlative light and electron microscopy of early *Caenorhabditis elegans* embryos in mitosis. *Meth. Cell Biol.* **79**, 99–117 (in the press).
26. Kuriyama, R. & Borisy, G. G. Centriole cycle in Chinese hamster ovary cells as determined by whole-mount electron microscopy. *J. Cell Biol.* **91**, 814–821 (1981).
27. Robbins, E., Jentsch, G. & Micali, A. The centriole cycle in synchronized HeLa cells. *J. Cell Biol.* **36**, 329–339 (1968).
28. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
29. Delattre, M., Canard, C. & Gönczy, P. Sequential protein recruitment of *C. elegans* centriole formation. *Curr. Biol.* **16**, 1844–1849 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Dammermann and K. Oegema for their gift of the GFP-ZYG-1 strain and for sharing unpublished results; F. Barr for anti-GFP antibodies; M. Tipsword, M. Ruer and D. Drechsel for generating the ZYG-1 antibodies; A. Pozniakovsky for generating the GFP-SAS-5 and GFP-SAS-6 constructs; O. Molodtsova for biolistic bombardment; J. Mäntler and Q. DeRobillard for help with electron microscopy; C. Lorenz for secretarial assistance; K. McDonald, P. Verkade and I. Baines for help in developing the single-embryo high-pressure freezing approach; C. Eckmann for providing the FOG-1 feeding construct; and F. Friedrich for help with graphical artwork. We are grateful to G. Warren, W. Zachariae, R. Kittler and F. Quittnat Pelletier for their comments on the manuscript, and Hyman laboratory members for stimulating discussions. The Boulder Laboratory for 3D Electron Microscopy of Cells is supported by a grant from the National Institutes of Health to J. R. McIntosh. L.P. is supported by a postdoctoral fellowship from the HFSP.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.P. (pelletie@mpi-cbg.de).

LETTERS

p63 protects the female germ line during meiotic arrest

Eun-Kyung Suh^{1*}, Annie Yang^{1,2*}, Arminja Kettenbach^{1*}, Casimir Bamberger¹, Ala H. Michaelis¹, Zhou Zhu³, Julia A. Elvin⁵, Roderick T. Bronson⁴, Christopher P. Crum⁵ & Frank McKeon¹

Meiosis in the female germ line of mammals is distinguished by a prolonged arrest in prophase of meiosis I between homologous chromosome recombination and ovulation¹. How DNA damage is detected in these arrested oocytes is poorly understood, but it is variably thought to involve p53, a central tumour suppressor in mammals^{2–4}. While the function of p53 in monitoring the genome of somatic cells is clear, a consensus for the importance of p53 for germ line integrity has yet to emerge. Here we show that the p53 homologue p63 (refs 5, 6), and specifically the TAp63 isoform, is constitutively expressed in female germ cells during meiotic arrest and is essential in a process of DNA damage-induced oocyte death not involving p53. We also show that DNA damage induces both the phosphorylation of p63 and its binding to p53 cognate DNA sites and that these events are linked to oocyte death. Our data support a model whereby p63 is the primordial member of the p53 family and acts in a conserved process of monitoring the integrity of the female germ line, whereas the functions of p53 are restricted to vertebrate somatic cells for tumour suppression. These findings have implications for understanding female germ line fidelity, the regulation of fertility and the evolution of tumour suppressor mechanisms.

The discovery of two p53-like genes, namely p63 and p73, in mammals has provoked much speculation about their individual and collective functions⁶. For instance, do p53, p63 and p73, which regulate many genes in common⁷ cooperate in tumour suppression and genome integrity? Gene knockout models in mice implicate p63 and p73 in epithelial stem cells⁸ and neurogenic processes⁹, respectively, but the complex pattern of transcription from these genes⁵ (Fig. 1a) implies a diverse set of isoforms with apparently contradictory functions. Thus, the p63 isoform $\Delta Np63\alpha$ is highly expressed in the stem cells of stratified epithelia and is required for their maintenance in a process that is independent of TAp63 (refs 5, 8, and F. Pinto, M. Senoo and F.M., unpublished observations). In contrast, the roles of the TAp63 isoforms have remained more enigmatic. To probe the function of TAp63 we generated monoclonal antibodies specific to the transactivation domain of TAp63 (Supplementary Fig. 1) and analysed late-stage murine embryos by immunohistochemistry. The ovary—and in particular oocyte nuclei—was the only site of strong expression of TAp63 (Fig. 1b). Analyses with additional p63 antibodies^{5,10} led us to conclude that the predominant p63 isoform in oocytes is TAp63 α (Supplementary Fig. 1a–f). Our TAp63-specific antibodies showed no detectable signal in testes of newborn or adult mice (Fig. 1c, d), indicating that germ line expression of TAp63 α (hereafter denoted TAp63) is biased towards the female.

We next examined whether TAp63 expression varied during oogenesis. This process starts with the migration of diploid mitotic oogonia to the genital ridge between embryonic day (E)12 and E13 (ref. 11). These cells develop into primary oocytes that progress through non-reductive DNA replication, homologous chromosome recombination between E13 and E18.5, and finally arrest in prophase of meiosis I (dictyate arrest) between E18.5 and five days after birth (P5)¹. Dictyate arrest is specific to meiosis in females pending recruitment of arrested oocytes for ovulation, and potentially lasts for more than a year in mice and for many decades in humans. We probed ovaries from embryos derived from timed pregnancies with TAp63 monoclonal antibodies as well as the oocyte marker Msy2 (ref. 12). At E16.5 we found no TAp63-positive oocytes, whereas at E18.5 about 20% expressed detectable TAp63 (Fig. 1e, f). Progressively, though, the percentage of positive cells increased from that point such that in pups at P5, when oocytes are uniformly in dictyate arrest, essentially all the oocytes showed strong TAp63 expression. TAp63 expression remained high in primordial and primary follicles but was lost in more advanced follicles being recruited for ovulation (Supplementary Fig. 2). Our analysis of p63-null⁸ ovaries indicated that oogenesis and folliculogenesis were similar to those observed in wild-type animals (Supplementary Fig. 3). Given its dispensability for oogenesis, we examined whether TAp63 has a function in the oocyte's response to DNA damage. In conformity with previous studies¹³, exposure of wild-type mice to 0.45 Gy of radiation resulted in a severe and specific loss of the small, primordial follicles from the ovary within five days (Fig. 2a). The larger, pre-antral oocytes survived these levels of ionizing radiation (Fig. 2a). Coincident with the loss of the small primary oocytes, TAp63 expression was absent from the ovary five days after irradiation (Fig. 2b). At shorter intervals after irradiation, we observed large numbers of oocytes that showed hypercondensed chromatin and TUNEL (TdT-mediated dUTP nick end labelling) staining, indicating that either apoptosis or necrosis¹⁴ was initiated within 24 h (Supplementary Fig. 4). We also noted that TAp63 underwent an electrophoretic mobility shift after exposure to ionizing radiation (Fig. 2c). By 12 h after irradiation, however, TAp63 was less evident, and by 16 h TAp63 was only marginally detectable. This mobility shift in TAp63 was sensitive to λ -phosphatase treatment (Supplementary Fig. 5).

p53 is important in the apoptotic response to DNA damage by a wide range of somatic cells^{2–4}. Because the involvement of p53 in germ line cell death remained controversial, we examined how oocytes in p53-deficient mice respond to ionizing radiation. Significantly, the pattern and extent of oocyte loss in p53-null mice was very similar to that of wild-type controls (Fig. 2d), arguing

¹Department of Cell Biology, ²Department of Biological Chemistry and Molecular Pharmacology, ³Department of Genetics and ⁴Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

against an essential role of p53 in DNA damage-induced death of these oocytes.

To test genetically whether TAp63 was required for DNA damage-induced oocyte death, we generated mice lacking exons 2 and 3 of the *p63* gene encoding the TA-specific amino terminus (Fig. 1a). These animals retain the $\Delta Np63$ isoforms, show normal epithelial morphogenesis, and are viable (A.K., A.Y., F.M. and F. Pinto, unpublished observations). Significantly, the oocytes in primordial follicles of the TAp63-null mice were resistant to the same dose of radiation that

killed virtually all those in wild-type and p53-null ovaries (Fig. 2e, f). Similar results were obtained when we compared wild-type or p53-null ovaries grown and irradiated *in vitro* with ovaries derived from p63-null mice (Supplementary Fig. 6). Together, these data indicate that TAp63, but not p53, is essential for DNA-damage-induced primary oocyte death.

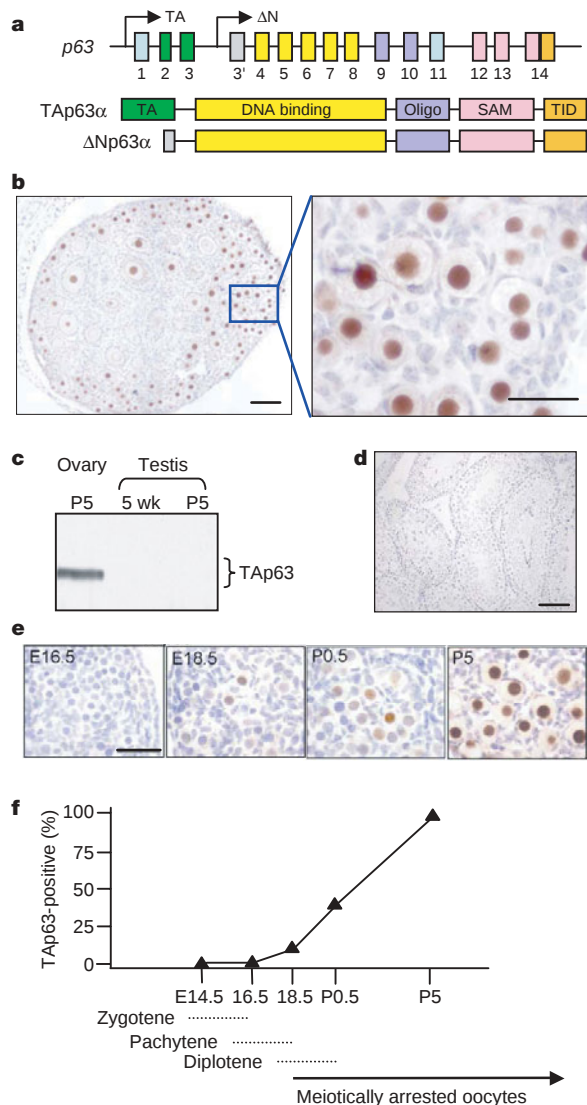


Figure 1 | TAp63 α protein is expressed in ovary but not testis. **a**, Diagram of the human *p63* gene at chromosome 3q27, showing the positions of the exons and major promoters TA, driving isoforms having a transactivation (TA) domain, and ΔN , controlling the expression of ΔN isoforms that lack the TA domain. Domains common to TAp63 α and $\Delta Np63\alpha$ include the DNA-binding domain, the oligomerization domain (oligo), the sterile alpha motif (SAM) and the transactivation inhibitory domain (TID). Primary sequence identity between human p63 and p53 is indicated. **b**, Expression of TAp63 in a section of ovary from P5 mouse, stained with a monoclonal antibody specific for the TA domain of TAp63. Scale bar, 50 μ m. **c**, Western blot against lysates from P5 ovaries and from testes of P5 and five-week-old (5 wk) animals with the use of an anti-TAp63 monoclonal antibody. **d**, Immunohistochemical analysis with anti-TAp63 monoclonal antibody of section of testis from five-week-old mouse. The blue staining is from haematoxylin dye. Scale bar, 100 μ m. **e**, Anti-TAp63 immunohistochemistry of staged ovaries. Scale bar, 25 μ m. **f**, Graphical representation of percentage of oocytes in **e** that stained positive with an anti-TAp63 monoclonal antibody, together with the corresponding phases of meiosis I.

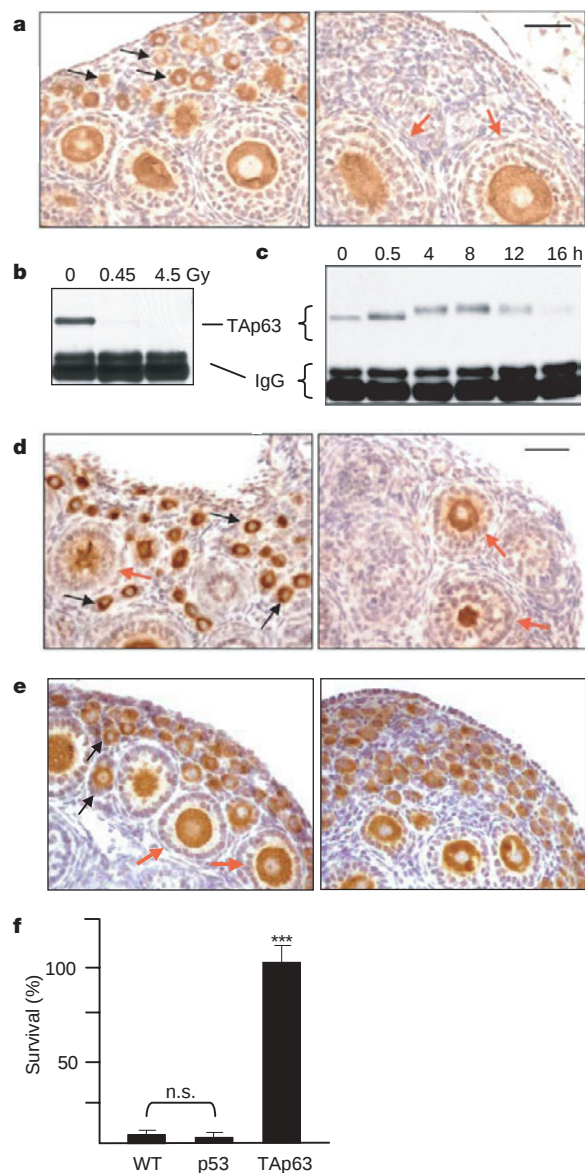


Figure 2 | Oocyte death after DNA damage requires p63. **a**, Anti-Msy2 immunohistochemistry of ovaries of P10 wild-type mice five days after receiving 0 Gy (left) and 0.45 Gy (right) of ionizing radiation. Black arrows indicate the smaller, more numerous oocytes in primordial follicles; red arrows indicate the more advanced, preantral follicles. Scale bar, 50 μ m. **b**, Western blot of wild-type ovaries from mice irradiated as indicated at P5 and killed at P10; stained with TAp63-specific antibodies. The immunoglobulin (IgG) signal is from endogenous sources. **c**, Western blot of lysates of wild-type ovaries from P5 mice taken at the indicated times after receiving 4.5 Gy ionizing radiation. **d**, Analysis of response of p53-null ovaries to ionizing radiation. Anti-Msy2 immunohistochemistry. Mice received 0 Gy (left) and 0.45 Gy (right) of ionizing radiation at P5 and were examined at P10. Scale bar, 50 μ m. **e**, Immunohistochemistry with antibodies against Msy2 on sections of P10 ovary of TAp63-null mice taken five days after either 0 Gy (left) or 0.45 Gy (right) ionizing radiation. Scale as in **d**. **f**, Quantification of the survival profiles of wild-type, p53-null and TAp63-null oocytes from primordial follicles five days after irradiation at 0.45 Gy. Error bars show s.d.; $n \geq 3$. Three asterisks, $P < 0.001$ compared with wild type (WT) and p53 $^{-/-}$; n.s., not significant.

We next examined whether there was a threshold of DNA damage that induced oocyte death and how this might be related to the observed mobility shift in TAp63. We exposed P5 wild-type mice to a spectrum of doses of ionizing radiation and analysed their ovaries over the next 24 h. In accordance with previous observations¹³, the threshold for death of oocytes in primordial follicles seemed remarkably sensitive to the precise amount of radiation: most oocytes survive a dose of 0.1 Gy, whereas virtually all die at 0.45 Gy (Fig. 3a). Using antibodies against a phospho-epitope on histone 2AX (γ -H2AX) to mark sites of double-strand break repair¹⁵, we determined that oocytes receiving 0.1 Gy sustained, on average, three double-strand breaks, whereas those exposed to 0.45 Gy had about ten double-strand breaks (Fig. 3a, b, and Supplementary Fig. 7). These findings indicate that, as with the activation of p53 in somatic cells¹⁶, the threshold for the cell death response is probably determined by one or a few irreparable DNA breaks in oocytes.

Given the tight dose–response relationship between radiation and oocyte death, we examined whether there was also a correlation with

TAp63 phosphorylation. TAp63 mobility was followed by western blot analysis at progressive intervals after defined radiation doses (Fig. 3c). At the higher doses chosen for this experiment, 0.3 and 0.45 Gy, virtually all of the TAp63 underwent a mobility shift at 8 h, and by 24 h little or no TAp63 was detectable in these ovaries. Because TAp63 disappears upon oocyte cell death, the loss of TAp63 after doses of 0.3 and 0.45 Gy probably reflects the loss of oocytes at 24 h. However, at low levels of radiation, particularly 0.1 Gy, only a fraction of the TAp63 underwent a mobility shift and much TAp63 remained at 24 h. Consistently, large numbers of oocytes survived irradiation at 0.1 Gy even after five days, whereas those receiving 0.45 Gy showed almost no primordial follicles (Fig. 3a). Thus, there seemed to be a relationship between the fraction of TAp63 that undergoes a mobility shift and the fraction of oocytes that die in response to DNA damage. Finally, we examined how the kinetics of the TAp63 mobility shift influenced the kinetics of oocyte loss. TAp63 underwent a more rapid mobility shift after receiving 4.5 Gy radiation than it did after 0.45 Gy (Fig. 3d). Accordingly, we found

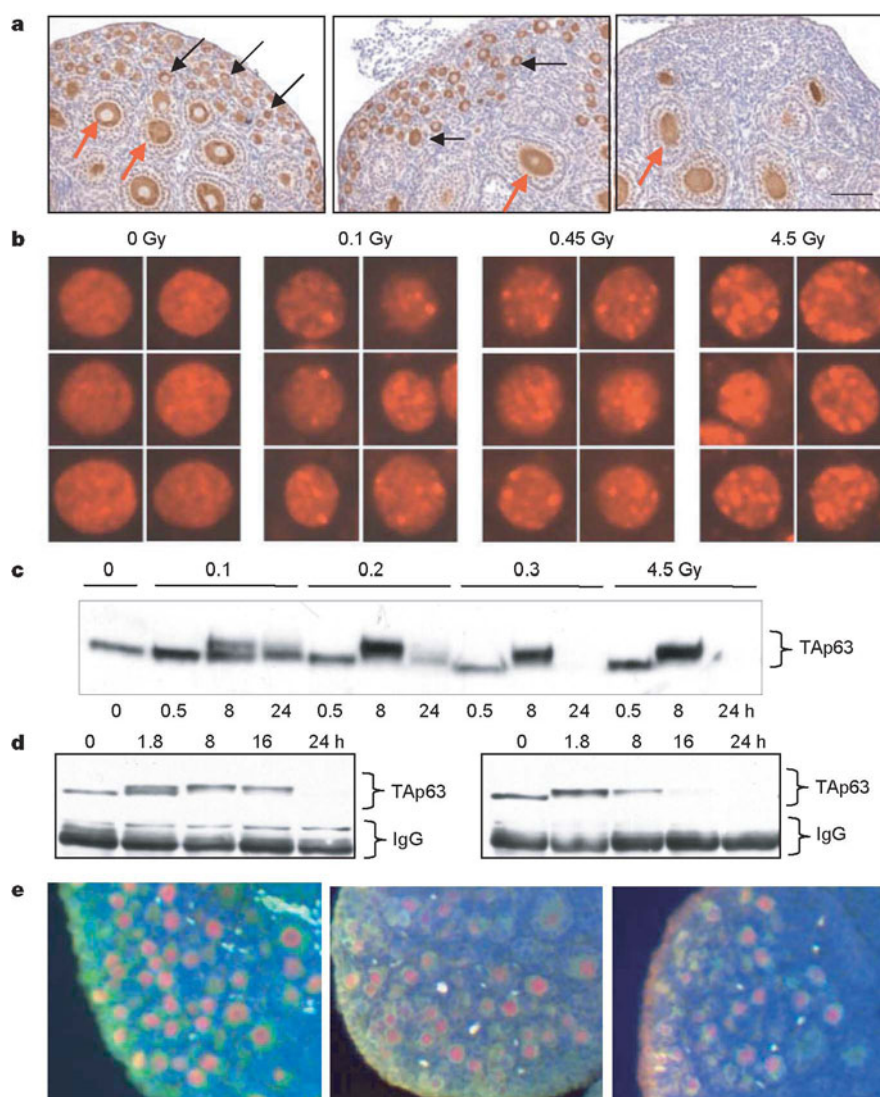


Figure 3 | Threshold for TAp63 modifications and oocyte cell death.

a, Anti-Msy2 immunohistochemistry on sections of P10 ovary from mice irradiated at P5 with 0 Gy (left), 0.1 Gy (middle) and 0.45 Gy (right) ionizing radiation. Black and red arrows indicate oocytes in primordial and preantral follicles, respectively. Scale bar, 50 μ m. **b**, Immunofluorescence analysis with anti- γ H2AX antibodies to reveal sites of DNA double-strand breaks in primary oocytes 3 h after exposure to the indicated doses of ionizing radiation. **c**, Western analysis of TAp63 in ovaries from wild-type mice at P5

after ionizing radiation at the indicated doses and the indicated post-radiation intervals. **d**, Western blots of lysates of ovaries of five-day-old mice obtained at the indicated times after irradiation with 0.45 Gy (left) and 4.5 Gy (right) ionizing radiation, probed with anti-TAp63 antibodies. **e**, Immunofluorescence images obtained with antibodies against TAp63 (red) and the oocyte marker Msy2 (green) on sections through ovaries of five-day-old wild-type mouse pups 12 h after treatment with 0 Gy (left), 0.45 Gy (middle) and 4.5 Gy (right) ionizing radiation.

that oocytes receiving 4.5 Gy radiation progressed to cell death sooner than those receiving only 0.45 Gy (Fig. 3e). Together, these data suggest a temporal link between the onset of DNA-damage-dependent Tap63 phosphorylation and the death of oocytes in primordial follicles.

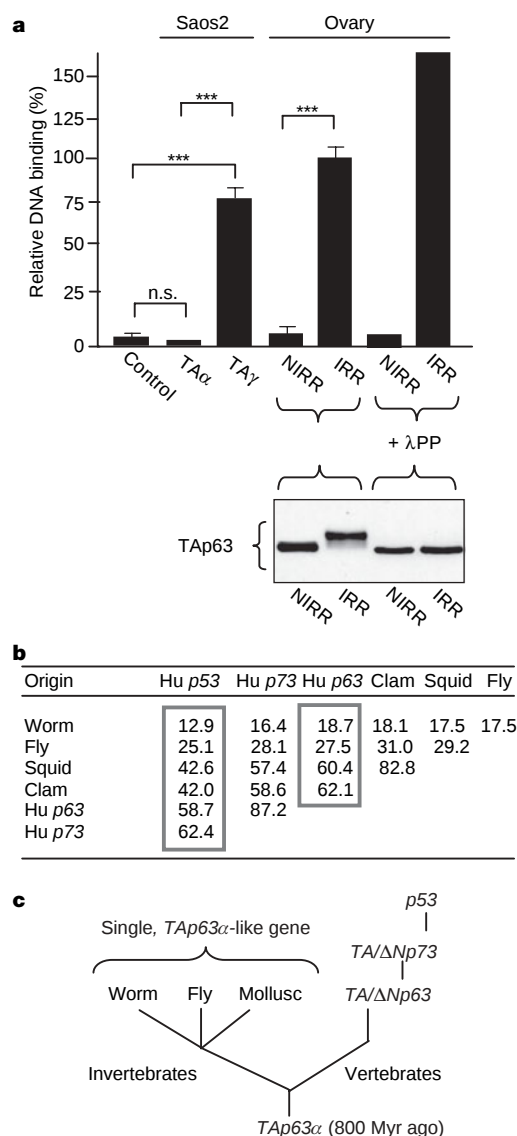


Figure 4 | Tap63: DNA-damage-dependent DNA binding and evolution.

a, Top: histogram of relative DNA binding of Tap63α (TAα) and Tap63γ (TAγ) derived from transfected Saos2 cells, Tap63 from non-irradiated (NIRR) and irradiated with 4.5 Gy (IRR) wild-type ovaries, and Tap63 from non-irradiated and irradiated ovaries after treatment with λ protein phosphatase (λPP). Error bars show s.d.; $n \geq 3$. Three asterisks, $P < 0.001$ versus Tap63γ or versus non-irradiated ovary; n.s., not significant. Bottom, western blot of lysates used in the corresponding histogram, probed with anti-Tap63 antibodies. **b**, Comparison of percentage identity within the DNA-binding domain of p53-related genes in invertebrates (nematode worm *Caenorhabditis elegans*, fruitfly *Drosophila melanogaster*, northern European squid *Loligo forbesi* and softshell clam *Mya arenaria*) and the human (Hu) p53, p63 and p73 genes. Primary sequence identity across these species is indicated, with those to human (Hu) p53 and human p63 enclosed by boxes. **c**, Overall deduced phylogeny of the p53 family of transcription factors. Tap63α is depicted as the primordial gene present at the split in Bilateria and the only one present in invertebrates. In the rise of vertebrates, the Tap63 gene gained an internal promoter to produce ΔN isoforms. This two-promoter gene was duplicated in vertebrates to produce the p73 gene, which then duplicated to give rise to p53.

The association between Tap63 phosphorylation and oocyte cell death suggested that Tap63 is somehow activated in response to DNA damage. The basis for the transcriptional inactivity of Tap63α in standard p53 reporter assays is incompletely understood⁵, but the most compelling model is based on the discovery of a carboxy-terminal transactivation inhibitory domain that interacts directly with the N-terminal transactivation domain¹⁷. We therefore examined whether there was a difference in DNA-binding abilities between Tap63α and Tap63γ, the latter of which lacks the transactivation inhibitory domain and shows high transactivation and proapoptotic activity in transfection models^{5,17}. Tap63γ derived from transfected Saos-2 cells showed a 25-fold increase in DNA binding over that of Tap63α, suggesting that the latter was functionally silent because of inefficient DNA binding (Fig. 4a). We next compared Tap63 from non-irradiated and irradiated ovaries with the use of the same DNA binding assay. Significantly, phosphorylated Tap63 showed an increase in DNA binding to nearly 20-fold that of the non-irradiated oocytes, indicating a DNA-damage-dependent disinhibition of DNA binding (Fig. 4a, and Supplementary Fig. 7). However, treatment of the hyperphosphorylated Tap63 in lysates from irradiated oocytes with λ protein phosphatase did not reverse its DNA-binding capability and, if anything, seemed to enhance this ability (Fig. 4a). We speculate that the failure of the λ phosphatase to prevent the DNA binding of Tap63 was due to an inability to remove certain phosphates critical to Tap63 activation, or that DNA-damage-dependent hyperphosphorylation disrupts an autoinhibitory mechanism in Tap63 that is not readily restored by dephosphorylation. The molecular basis for this activation step, and whether alternative processes can stimulate Tap63 activity, represents an intriguing structural challenge.

The requirement for p63 in mammalian oocyte death after DNA damage is consistent with data derived from genetically tractable organisms such as flies and worms, each of which has a single p53-like gene^{18–25}. Although it has been generally assumed that these invertebrate genes are orthologues of the p53 gene, our phylogenetic analyses indicate that they are more related to the vertebrate p63 gene than they are to p53 (Fig. 4b). Moreover, these data indicate that Tap63 was the primordial member of the p53 family and that the p53 gene itself arose only during vertebrate evolution for tumour suppression (Fig. 4c).

Our data implicate Tap63 in maintaining the fidelity of the female germ line. The restriction of Tap63 function to the female germ line could reflect the sexual dimorphism in strategies of meiosis and gamete production²⁶. In particular, oocytes exist as a limited population largely arrested in a potentially vulnerable, tetraploid state for an extraordinary duration. A mechanism of surveying DNA damage in the oocyte would therefore seem essential to ensure the fidelity of the genome in the subsequent generation. Given that point mutations underlying *de novo* human genetic disease arise disproportionately from the male germ line²⁷, it will be important to define Tap63's role in this apparent bias. Our data also underscore an evolutionarily conserved relationship between DNA damage and activation of the p53 family of transcription factors^{28,29}. Specifically, we propose a model whereby Tap63 is the primordial p53 family member acting to maintain the female germ line, whereas the vertebrate-specific p53 gene functions only in somatic cells. These findings further demonstrate a novel function for p63 that has important medical implications for the loss of oocytes resulting from chemotherapy and other toxic stresses. Finally, the functional significance of prolonged meiotic arrest of oocytes has received scant attention despite its high conservation. How Tap63 is integrated into meiosis will probably reveal much about how this state of arrest affects oocyte fate, female fertility and, ultimately, germ line integrity.

METHODS

Mice. The p63^{-/-} (ref. 8) and p53^{-/-} Balb/c mice were described previously and genotyped with standard procedures. Tap63-deficient mice were generated by

replacing exons 2 and 3 with a neomycin resistance cassette in embryonic stem cells (A.K., A.Y. and F.M., unpublished observations).

Irradiation. P5 or P6 mice received a single dose of 0.1, 0.45 or 4.5 Gy of ionizing radiation on a rotating turntable in a ^{137}Cs irradiator (dose rate 141 rad min^{-1}) before being killed and having their tissue harvested at the indicated time points.

Anti-TAp63 monoclonal antibodies. Mice were immunized with a fusion protein of glutathione S-transferase with the first 69 amino acid residues of murine TAp63 (ref. 5). Spleen cells were fused with NSO myeloma cells and the antibodies from the resulting hybridomas were screened by immunofluorescence and western blotting against BHK cells transfected with expression plasmids driving murine TAp63 α .

Immunohistochemistry and quantification. Dissected ovaries were processed and stained as described in Supplementary Information. The percentages of p63-expressing oocytes at various developmental ages were derived from p63-positive oocyte counts relative to Msy2 counts in adjacent mid-sections of the ovary that included the hilus. In irradiation experiments, counts of Msy2-positive primordial and primary follicles were made from mid-ovary sections. The percentage survival of oocytes in the irradiated ovaries was calculated relative to the non-irradiated sister ovaries. *P* values were determined with the unpaired Student's *t*-test.

Ovary processing, λ -phosphatase treatment and immunoblotting. Ovaries from non-irradiated or whole-body γ -irradiated mice were homogenized, treated with λ -phosphatase, separated by SDS-PAGE and probed with anti-TAp63 antibodies as described in Supplementary Information.

DNA binding assay. Lysates from control and irradiated (4.5 Gy, isolated 2 h after dose delivery) ovaries were incubated in the presence of $0.5\text{ }\mu\text{g}$ of poly(dI)•poly(dC) with 2 pmol of biotinylated double-strand oligonucleotide representing the p53 DNA binding consensus, mutant versions thereof⁶⁰ or randomized sequence. Bound TAp63 α was detected with TAp63-specific monoclonal antibodies and quantified with the TransFactor Universal Chemiluminescence Detection Kit (Clontech) in accordance with the manufacturer's recommendations. Oligonucleotide sequences and additional details are provided in Supplementary Information.

Received 5 July; accepted 12 October 2006.

Published online 22 November 2006.

- Eppig, J. J. Coordination of nuclear and cytoplasmic oocyte maturation in Eutherian mammals. *Reprod. Fertil. Dev.* **8**, 485–489 (1996).
- Hainaut, P. & Wiman, K. *25 Years of p53 Research* (Springer, Berlin, 2005).
- Lane, D. P. *et al.* On the regulation of the p53 tumour suppressor, and its role in the cellular response to DNA damage. *Phil. Trans. R. Soc. Lond. B* **347**, 83–87 (1995).
- Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
- Yang, A. *et al.* p63, a p53 homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **2**, 305–316 (1998).
- Yang, A., Kaghad, M., Caput, D. & McKeon, F. On the shoulders of giants: p63, p73 and the rise of p53. *Trends Genet.* **18**, 90–95 (2002).
- Yang, A. *et al.* Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol. Cell* (in the press).
- Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
- Yang, A. *et al.* p73-deficient mice have neurological, phenomonal and inflammatory defects but lack spontaneous tumours. *Nature* **404**, 99–103 (2000).
- Di Iorio, E. *et al.* Isoforms of ΔNp63 and the migration of ocular limbal cells in human corneal regeneration. *Proc. Natl Acad. Sci. USA* **102**, 9523–9528 (2005).
- Peters, H. Migration of gonocytes into the mammalian gonad and their differentiation. *Phil. Trans. R. Soc. Lond. B* **259**, 91–101 (1970).
- Yu, J., Hecht, N. B. & Schultz, R. M. Expression of MSY2 in mouse oocytes and preimplantation embryos. *Biol. Reprod.* **65**, 1260–1270 (2001).

- Peters, H. & Levy, E. Effect of irradiation in infancy on the mouse ovary: a quantitative study of oocyte sensitivity. *J. Reprod. Fertil.* **7**, 37–45 (1964).
- Budihardjo, I., Oliver, H., Lutter, M., Luo, X. & Wang, X. Biochemical pathways of caspase activation during apoptosis. *Annu. Rev. Cell Dev. Biol.* **15**, 269–290 (1999).
- Mahadevaiah, S. K. *et al.* Recombinational DNA double-strand breaks in mice precede synapsis. *Nature Genet.* **27**, 271–276 (2001).
- Ma, L. *et al.* A plausible model for the digital response of p53 to DNA damage. *Proc. Natl Acad. Sci. USA* **102**, 14266–14271 (2005).
- Serber, Z. *et al.* A C-terminal inhibitory domain controls the activity of p63 by an intramolecular mechanism. *Mol. Cell. Biol.* **22**, 8601–8611 (2002).
- Derry, W. B., Putzke, A. P. & Rothman, J. H. *Caenorhabditis elegans* p53: role in apoptosis, meiosis, and stress resistance. *Science* **294**, 591–595 (2001).
- Schumacher, B., Hofmann, K., Boulton, S. & Gartner, A. The *C. elegans* homolog of the p53 tumor suppressor is required for DNA damage-induced apoptosis. *Curr. Biol.* **11**, 1722–1727 (2001).
- Ollmann, M. *et al.* *Drosophila* p53 is a structural and functional homolog of the tumor suppressor p53. *Cell* **101**, 91–101 (2000).
- Brodsky, M. H. *et al.* *Drosophila* p53 binds a damage response element at the reaper locus. *Cell* **101**, 103–113 (2000).
- Jin, S. *et al.* Identification and characterization of a p53 homologue in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 7301–7306 (2000).
- Sogame, N., Kim, M. & Abrams, J. M. *Drosophila* p53 preserves genomic stability by regulating cell death. *Proc. Natl Acad. Sci. USA* **100**, 4696–4701 (2003).
- Brodsky, M. H. *et al.* *Drosophila melanogaster* MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. *Mol. Cell. Biol.* **24**, 1219–1231 (2004).
- Schumacher, B. *et al.* Translational repression of *C. elegans* p53 by GLD-1 regulates DNA damage-induced apoptosis. *Cell* **120**, 357–368 (2005).
- Hunt, P. A. & Hassold, T. J. Sex matters in meiosis. *Science* **296**, 2181–2183 (2002).
- Haldane, J. B. S. The rate of mutation of the gene for hemophilia and its segregation in males and females. *Ann. Eugen.* **13**, 262–271 (1947).
- Ko, L. J. & Prives, C. p53: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
- Bourdon, J. C. *et al.* p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* **19**, 2122–2137 (2005).
- el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nature Genet.* **1**, 45–49 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Schultz, B. Spyropoulos, P. Moens and M. De Luca for gifts of antibodies; A. Lee, G. Perez, C. Morton and Y. Lee for advice and assistance with technical procedures; and V. Doetsch, M. Senoo, F. Pinto, H. Green, A. Sharpe and M. Colaiacova for discussions. A.Y. was supported by a fellowship from the Department of Defence Program in Breast Cancer. This work was supported by grants from the NCI and the NIH to C.P.C. and F.M.

Author Contributions E.K.S. established the ovary experimental system, performed the DNA damage experiments with the wild-type, p53-null, and p63-null mice *in vivo* and *in vitro*, identified the TAp63 phosphorylation-dependent mobility shift and performed the kinetics linking it to oocyte death. A.Y. generated the anti-TAp63 monoclonal antibodies, designed the TAp63 targeting and genotyping strategies and assisted in the generation of the TAp63 knockout mice. A.K. generated the TAp63 knockout mice and performed the DNA damage experiments on these strains. C.B. performed the TAp63 DNA binding experiments and statistical analyses of data presented in this paper. A.M. generated and characterized the anti-TAp63 antibodies with A.Y. Z.Z. and A.K. performed the bioinformatics analysis of the p53 gene family. J.E. made early contributions on TAp63 expression in human oocytes. R.B. supervised the histopathology analysis. C.C. and F.M. directed the project, and F.M. wrote the manuscript with help from A.Y., A.K. and E.K.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.M. (fmckeon@hms.harvard.edu).

Repression of p53 activity by Smyd2-mediated methylation

Jing Huang^{1*}, Laura Perez-Burgos^{2*}, Brandon J. Placek¹, Roopsha Sengupta², Mario Richter², Jean A. Dorsey¹, Stefan Kubicek², Susanne Opravil², Thomas Jenuwein² & Shelley L. Berger¹

Specific sites of lysine methylation on histones correlate with either activation or repression of transcription^{1–3}. The tumour suppressor p53 (refs 4–7) is one of only a few non-histone proteins known to be regulated by lysine methylation⁸. Here we report a lysine methyltransferase, Smyd2, that methylates a previously unidentified site, Lys 370, in p53. This methylation site, in contrast to the known site Lys 372, is repressing to p53-mediated transcriptional regulation. Smyd2 helps to maintain low concentrations of promoter-associated p53. We show that reducing Smyd2 concentrations by short interfering RNA enhances p53-mediated apoptosis. We find that Set9-mediated methylation of Lys 372 inhibits Smyd2-mediated methylation of Lys 370, providing regulatory cross-talk between post-translational modifications. In addition, we show that the inhibitory effect of Lys 372 methylation on Lys 370 methylation is caused, in part, by blocking the interaction between p53 and Smyd2. Thus, similar to histones, p53 is subject to both activating and repressing lysine methylation. Our results also predict that Smyd2 may function as a putative oncogene by methylating p53 and repressing its tumour suppressive function.

Members of the Smyd (SET and MYND domain) family and Suv4-20 were expressed in bacteria and tested for histone and p53 methylation. Smyd2, but not Smyd3, Smyd5 or Suv4-20, methylated a peptide spanning amino acid residues 358–393 in the carboxy terminus of p53 (Supplementary Fig. S1a). Peptide mapping showed that lysine 370 (K370) is the methylation site for Smyd2 (Fig. 1a, b). Unmodified, mono-, di- and tri-methylated K370 p53 peptides (residues 361–380) were subjected to Smyd2 methylation *in vitro* followed by mass spectrometry (Supplementary Fig. S1b). The results showed that Smyd2 mono-methylates K370 in p53 *in vitro*.

Polyclonal antibodies specifically recognizing mono-methylated (p53K370me1), di-methylated (p53K370me2) and tri-methylated (p53K370me3) K370 were generated and tested for specificity *in vitro*

(Supplementary Fig. S2a). These antibodies were used to test p53 methylation by Smyd2 *in vivo*. H1299 cells were transfected with Flag-p53, Flag-p53(K370A) or Flag-p53(K370R) (Fig. 2a and Supplementary Fig. S2b), either with or without Flag-Smyd2 or a methylation-defective mutant Flag-Smyd2(MD) (Supplementary Fig. S3). Flag immunoprecipitates were subjected to western blot analysis with methylation-specific antibodies. Flag-p53, but not the K370R mutant, was detected by the antibody to p53K370me1 (Fig. 2a, compare lanes 1 and 4). The extent of K370 methylation of Flag-p53, but not the K370R mutant, was greatly increased by cotransfection with Flag-Smyd2 (Fig. 2a, compare lanes 2 and 5). Detection of p53 by the antibody to p53K370me1 was greatly diminished by

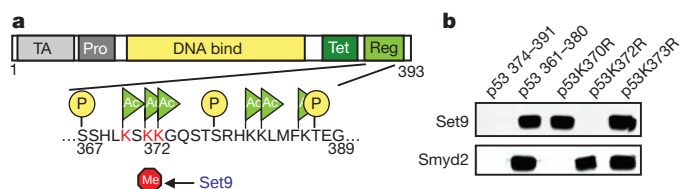


Figure 1 | Smyd2 methylates p53 at K370 *in vitro*. **a**, p53 functional domains. TA, transactivation; Pro, proline-rich; Tet, tetramerization; Reg, regulatory; Ac, acetylation; P, phosphorylation; Me, methylation. Set9 methylates p53 at K372. **b**, Autoradiogram of HMTase assay with recombinant Set9 (top) or Smyd2 (bottom) on p53 peptide 374–391, 361–380 or 361–380 bearing K370R, K372R or K373R substitutions.

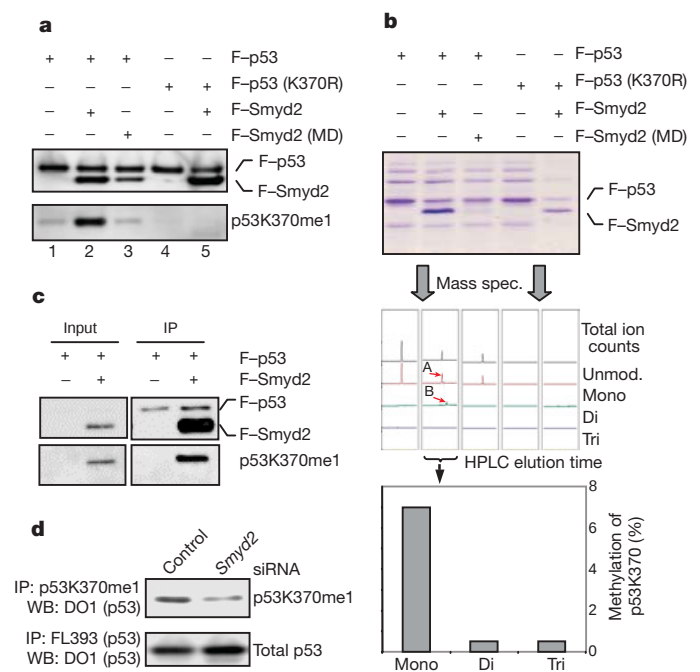


Figure 2 | Smyd2 methylates p53 at K370 *in vivo*. **a**, Western blot analysis of Flag immunoprecipitates from H1299 cells expressing Flag-p53 (lanes 1–3) or Flag-p53(K370R) (lanes 4–5) either alone (lanes 1 and 4) or with Flag-Smyd2 (lanes 2 and 5) or Flag-Smyd2(MD) (lane 3). **b**, Coomassie staining (top) and mass spectrometry analysis (middle and bottom) of Flag immunoprecipitates from H1299 cells. **c**, Western blot analysis of input and Flag immunoprecipitates from U2OS cells transfected with control or Smyd2 siRNA, followed by immunoprecipitation with antibody to p53K370me1 (top) or to full-length p53 (FL393; bottom). **d**, Western blot analysis of U2OS cells transfected with control or Smyd2 siRNA, followed by immunoprecipitation with antibody to p53K370me1 (top) or to full-length p53 (FL393; bottom).

¹Gene Expression and Regulation Program, The Wistar Institute, Philadelphia, Pennsylvania, 19104, USA. ²Research Institute of Molecular Pathology (IMP), The Vienna Biocenter, Vienna, A-1030, Austria.

*These authors contributed equally to this work.

cotransfection with Flag-Smyd2 (MD) (Fig. 2a, compare lanes 2 and 3). The antibody to p53K370me2 recognized p53 largely without regard to its methylation status at K370, whereas the antibody to p53K370me3 did not detect any tri-methylation signal (data not shown).

The methylation status of Flag-p53 in H1299 cells was also tested by mass spectrometry (Fig. 2b). Flag-p53 was immunoprecipitated and gel-isolated after Coomassie blue staining. The results showed that p53 is mono-methylated by Smyd2 *in vivo* (Fig. 2b).

We analysed methylation of p53 in U2OS cells. Cotransfected Flag-p53 and Flag-Smyd2 yielded mono-methylated p53 (Fig. 2c), but not di- or tri-methylated p53 (data not shown). The ability of endogenous Smyd2 to methylate endogenous p53 was tested. The amount of p53K370me1 was reduced when cells were treated with a

short interfering RNA (siRNA) targeting *Smyd2* as compared with a control siRNA (Fig. 2d; see Supplementary Fig. S4b, c, for siRNA controls). These *in vivo* data support our *in vitro* findings that Smyd2 mono-methylates p53 at K370.

To address the role of Smyd2 in regulating the function of p53, we used the human fibroblast cell line BJ-DNP53, which stably expresses a dominant-negative variant of p53 to inactivate the function of endogenous p53 (ref. 9). Reducing Smyd2 concentration with *Smyd2* siRNA correlated with increased *p21* and *mdm2* expression in control BJ cells but not in BJ-DNP53 cells (Fig. 3a and Supplementary Fig. S4a) with or without irradiation. Smyd2 reduction also resulted in an increase in mRNA (Supplementary Fig. S4b) and protein (Supplementary Fig. S4c) concentrations of p21 and mdm2 in U2OS cells. However, this effect was muted in H1299 cells, which are p53 null (Supplementary Fig. S4c). Taken together, these results show that Smyd2 negatively regulates p53-responsive genes in a p53-dependent manner.

To determine whether Smyd2 regulates the function of p53 specifically through K370 methylation, wild-type p53, p53(K370R) or an empty expression vector was transfected into H1299 cells stably expressing control or *Smyd2* short hairpin RNA (shRNA; Fig. 3b). Wild-type p53 induced p21 expression to a greater extent in cells expressing *Smyd2* shRNA than in cells expressing control shRNA. However, p53(K370R) showed the same transcriptional ability in both cell lines, and this ability matched the higher activity of wild-type p53 in H1299 cells expressing *Smyd2* shRNA (Fig. 3b). These findings indicate that Smyd2 may downregulate the transcriptional activation ability of p53 through K370 methylation.

We investigated whether the p53K370me1 modification occurs on p53 associated with cognate genes. Double chromatin immunoprecipitation (ChIP) was done with Flag antibody for the first ChIP, followed by Flag peptide elution, and then p53K370me1 antibody for the second ChIP. The total p53 signal decreased when Smyd2 was cotransfected, whereas the p53K370me1 ChIP essentially remained the same (data not shown). Thus, the ratio of p53K370me1 signal to total p53 signal increased in response to cotransfected Smyd2 (Fig. 3c, compare lanes 1 and 2). This increase was eliminated in the *Smyd2*(MD) mutant (Fig. 3c, compare lanes 2 and 3). The higher ChIP signal required intact K370, because the increase was not seen in Flag-p53(K370R) immunoprecipitates (Fig. 3c, compare lanes 1–3 and lanes 4–6). The higher signal of p53K370me1, as compared with IgG, arising from p53(K370R) immunoprecipitates presumably results from the residual recognition of unmodified p53 by the methylation antibody when p53 is overexpressed (Fig. 3c, lanes 4–6).

The endogenous amounts of total p53 and K370-methylated p53 bound to the p21 promoter were assessed by ChIP assay in U2OS cells (Supplementary Fig. S5a). Although the promoter-associated amount of total p53 increased markedly on adriamycin treatment, K370-methylated p53 increased only slightly (Supplementary Fig. S5a), resulting in a decrease in the relative percentage of K370-methylated p53 associated with the p21 promoter (Supplementary Fig. S5b). These data further support the view that Smyd2-mediated K370 methylation is a repressive modification of p53.

We investigated the mechanism underlying the repressive effect of Smyd2-mediated K370 methylation. We found no change in the distribution of p53 between the nucleus and cytoplasm when Smyd2 was reduced (Supplementary Fig. S6). In addition, we detected no change in the amount of total p53 (Supplementary Figs S4c, S7 and S8d), K382-acetylated p53 and S15-phosphorylated p53 (Supplementary Fig. S7) when Smyd2 was reduced or increased.

Next, we examined the effect of increasing Smyd2 concentrations on p53 binding at the *p21* gene by ChIP using U2OS cell lines inducibly expressing either empty vector or Smyd2 (Supplementary Fig. S8a). Increasing Smyd2 lowered the amount of p53 bound to the *p21* promoter (Supplementary Fig. S8b). The decreasing amount of promoter-associated p53 correlated with decreasing amounts of *p21* mRNA (Supplementary Fig. S8c) and p21 protein (Supplementary

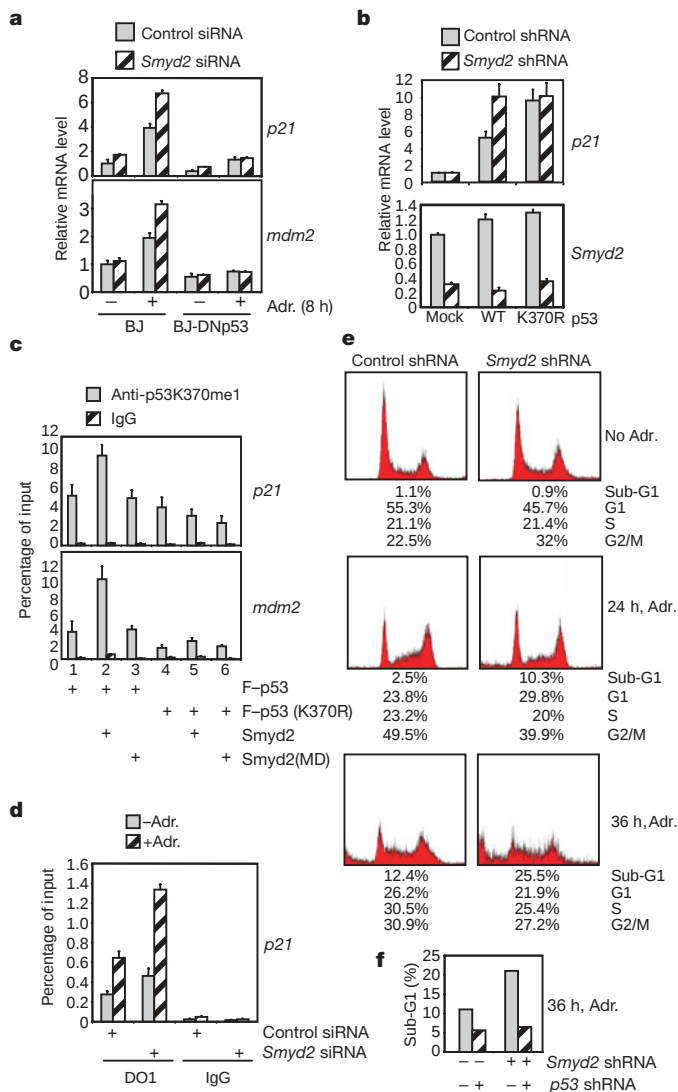


Figure 3 | Smyd2 represses the function of p53. **a**, Real-time PCR analysis of *p21* and *mdm2* mRNA in BJ and BJ-DNP53 cells transfected with control or *Smyd2* siRNA by mock or 8-h adriamycin treatment. **b**, Real-time PCR analysis of relative *Smyd2* and *p21* mRNA in H1299 cells expressing control shRNA or *Smyd2* shRNA and transfected with an empty, p53 or p53(K370R) expression vector. **c**, Sequential ChIP assay (Flag ChIP followed by p53K370me1 ChIP) to assess recruitment of p53K370me1 to the promoters of p53-responsive genes in H1299 cells. **d**, ChIP assay of U2OS cells transfected with control or *Smyd2* siRNA followed by mock or adriamycin treatment. **e**, Flow cytometry analysis of U2OS cells stably expressing control or *Smyd2* shRNA, followed by adriamycin treatment. **f**, U2OS cells stably expressing control or *Smyd2* shRNA were transduced with lentivirus bearing luciferase control or *p53* shRNA. Data are the mean $\pm$ s.d. (**a–d**).

Fig. S8d). We also reduced *Smyd2* by siRNA and observed an increase in promoter-associated p53 (Fig. 3d). These data suggest that K370-methylation of p53 reduces the DNA-binding efficiency of p53.

To address the role of *Smyd2* in p53-mediated cell-cycle arrest and apoptosis, we carried out flow cytometry analysis in U2OS cells stably

expressing either control or *Smyd2* shRNA with or without adriamycin treatment (Fig. 3e). Reduction of *Smyd2* did not cause obvious changes in the apoptotic fraction of cells in U2OS cells under the non-DNA damage condition (Fig. 3e, top). However, 24 h after DNA damage, fivefold more cells expressing *Smyd2* shRNA (10.3%) entered apoptosis (sub-G1) than did cells expressing control shRNA (2.5%; Fig. 3e, middle). This difference remained after 36 h of adriamycin treatment (Fig. 3e, bottom).

To address whether the effect of *Smyd2* on apoptosis is p53 dependent, we reduced the amounts of p53 protein and p53 mRNA in U2OS cells stably expressing control or *Smyd2* shRNA (Supplementary Figs S9 and S4c). Ablation of p53 almost completely eliminated the difference in apoptosis between U2OS cells expressing control shRNA and those expressing *Smyd2* shRNA (Fig. 3f). Because reintroduction of p53 into H1299 is known to induce apoptosis¹⁰, we transfected an empty or p53 expression vector into H1299 cells expressing control or *Smyd2* shRNA cells and subjected them to γ -irradiation (Supplementary Fig. S10). We found ectopically expressed p53 caused modest cell apoptosis without DNA damage (Supplementary Fig. S10, top, compare lanes 3 and 4 to lanes 1 and 2). On DNA damage, the addition of p53 resulted in more apoptosis in H1299 cells expressing *Smyd2* shRNA than in those expressing control shRNA (Supplementary Fig. S10, bottom, compare lanes 4 and 3). We were unable to use this p53 overexpression approach in H1299 cells to test further the prediction that the K370R mutant would cause even more cell death, because the apoptosis caused by overexpressed p53 was already severe. Taken together, we conclude that *Smyd2* is involved in p53-dependent cell-cycle arrest and apoptosis.

We examined the possible interplay between the activating modification p53K372me1 (ref. 8) and the repressing modification p53K370me1. We first tested whether there is any mutual effect on enzyme activity by using peptides and recombinant enzymes. Methylation by Set9 was not altered on peptides that were mono- or di-methylated at K370 (Fig. 4a, top). We also tested the ability of Set9 to mono-, di- or tri-methylate p53 peptides at K372, an issue that has not been reported⁸. We found that Set9 can mono- and di-methylate p53 at K372 *in vitro* (Supplementary Fig. S10a), and that methylation by *Smyd2* is lowered on both K372me1 and K372me2 peptides as compared with unmodified peptides (Fig. 4a, bottom). Taken together, our *in vitro* data suggest a 'one-way' cross-talk, in which K372 methylation inhibits K370 methylation.

We therefore tested whether there is a mutual effect *in vivo*. Whereas higher concentrations of p53K372me1 resulted in greatly reduced amounts of p53K370me1, higher concentrations of p53K370me1 did not lower the basal amounts of p53K372me1 (Fig. 4b, top). Furthermore, even the great increase in p53K372me1 obtained by cotransfection with Set9 was not lowered by *Smyd2* cotransfection (Fig. 4b, bottom).

A trivial explanation for the observed Set9/K372me1 inhibition of p53K370me1 could be physical blocking of the p53K370me1 antibody by methylation at the neighbouring K372 residue. To test this possibility, we synthesized peptides bearing the double modification K370me1 and K372me1. We found that detection of K370me1 was comparable to that of the singly modified K370me1 peptide (Supplementary Fig. S10b).

We also found that higher concentrations of p53K370me1 occur when Set9 is decreased by siRNA in H1299 cells that ectopically express p53 (Fig. 4c). Finally, we found that lowering endogenous Set9 increases endogenous amounts of p53K370me1 in U2OS cells (Fig. 4d). On the basis of the *in vitro* and *in vivo* results, we conclude that *Smyd2*-mediated methylation of K370 is inhibited by Set9-mediated methylation of K372.

We considered that the inhibitory effect of p53K372me1 on p53K370me1 might be caused by blocking the interaction between p53 and *Smyd2*. We found that Flag-p53 coprecipitated *Smyd2* in H1299 cells (Fig. 4e). Cotransfection of Set9 decreased the interaction

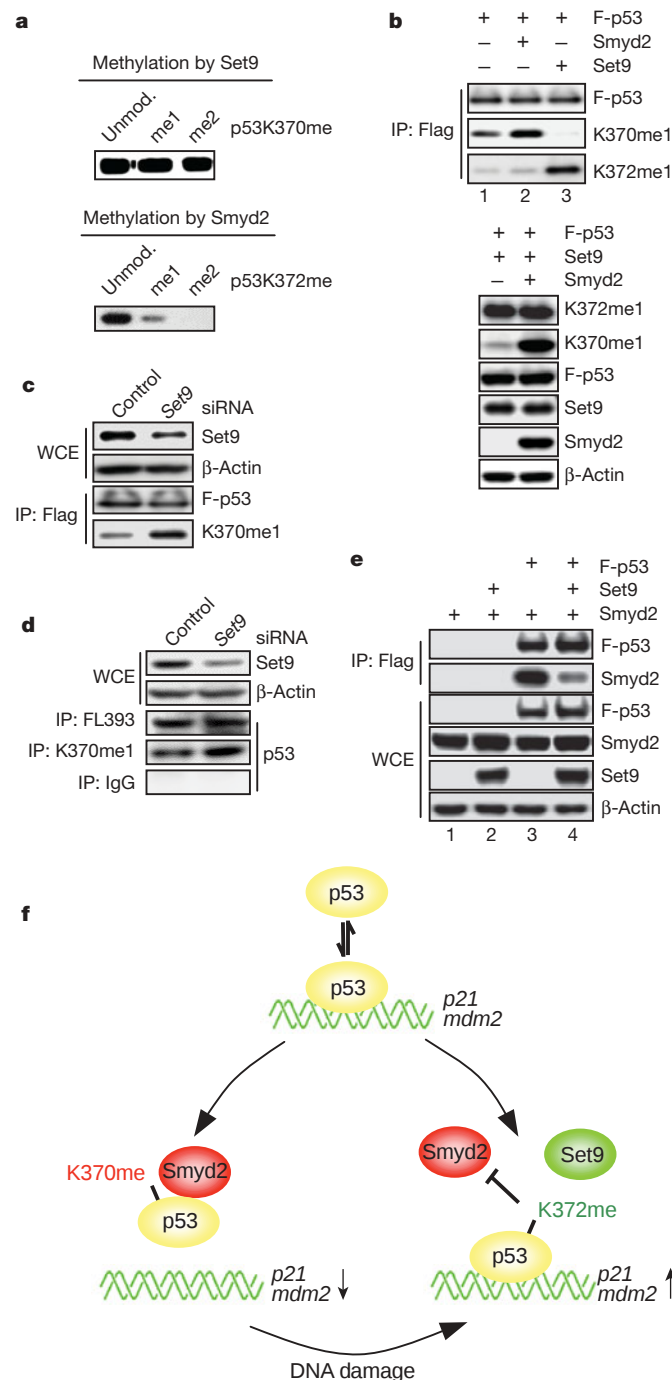


Figure 4 | Cross-talk between p53K370me1 and p53K372me1.

a, Autoradiogram of HMTase assays with recombinant Set9 or Smyd2 on p53 peptide 361–381. **b**, Western blot analysis of Flag immunoprecipitates (top) and whole-cell extract (bottom) from H1299 cells. **c**, Western blot analysis of whole-cell extract (WCE) and Flag immunoprecipitates from H1299 cells transfected with luciferase control or Set9 siRNA. **d**, Western blot analysis of whole-cell extract and immunoprecipitates obtained with the indicated antibodies from U2OS cells. **e**, Western blot analysis of whole-cell extract and Flag immunoprecipitates from H1299 cells expressing Smyd2 either alone (lane 1) or with Flag-p53 (lane 3), Set9 (lane 2), or both Flag-p53 and Set9 (lane 4). **f**, Model of the mechanism of Smyd2-mediated repression of p53.

between p53 and Smyd2 (Fig. 4e, compare lanes 3 and 4). Therefore, Set9 prevents Smyd2 from binding to p53.

Our results suggest that there is an equilibrium between promoter-bound and free p53 (Fig. 4f): Smyd2-mediated methylation of K370 shifts the equilibrium towards dissociation of p53 from DNA, whereas Set9-mediated methylation of K372 enhances the association of p53 with promoter by blocking Smyd2-mediated methylation of K370 and promotes activation of the *p21* and *mdm2* genes.

Histone modification cross-talk occurs, such that methylation of histone H3 on K9 (H3-K9) precludes H3-K4 methylation or phosphorylation of histone H3 on S10 (H3-S10)^{11,12}. We have shown here that the activating modification p53K372me1 causes interference to the repressing modification p53K370me1, which is reminiscent of the finding that Set9-mediated methylation of H3-K4 precludes Suv39h1-mediated methylation of H3-K9 (ref. 13). This non-proteolytic negative regulation through methylation of K370 may provide a pool of p53 protein to permit a quick response to DNA damage or cell stress.

Note added in proof: Smyd2 was recently shown to methylate histone H3 K36 (ref. 15).

METHODS

Histone methyltransferase assay. Histone methyltransferase (HMTase) assays were done as described⁸. In brief, 2 µg of peptide was incubated with 2 µg of recombinant enzyme and 1.1 µCi of *S*-adenosyl-methionine at 30 °C for 30 min.

Transfection. Cells were transfected with Lipofectamine 2000 (Invitrogen) for plasmid and with DharmaFECT 1 (Dharmacon) for siRNA (see Supplementary Information for details).

RT-PCR and real-time PCR. RT-PCR was done as described¹⁴ (see Supplementary Information for details).

ChIP and sequential ChIP assays. We carried out ChIP assays as described¹⁴. Real-time PCR was used to measure signals in input material and immunoprecipitates. The percentage of input was calculated as the immunoprecipitate signal over the input signal. Sequential ChIP assays are described in the Supplementary Information.

Lentivirus transduction. Lentiviral RNAi and expression systems were purchased from Invitrogen. shRNA oligonucleotide sequences were designed with BLOCK-iT RNAi Designer and are given in the Supplementary Information. Lentivirus bearing p53 shRNA was purchased from Sigma.

Received 21 July; accepted 22 September 2006.

Published online 15 November 2006.

1. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).

2. Lachner, M., O'Sullivan, R. J. & Jenuwein, T. An epigenetic road map for histone lysine methylation. *J. Cell Sci.* **116**, 2117–2124 (2003).
3. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
4. Olivier, M., Hussain, S. P., Caron de Fromental, C., Hainaut, P. & Harris, C. C. TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer. *IARC Sci. Publ.* **157**, 247–270 (2004).
5. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
6. Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
7. Prives, C. & Hall, P. A. The p53 pathway. *J. Pathol.* **187**, 112–126 (1999).
8. Chuikov, S. *et al.* Regulation of p53 activity through lysine methylation. *Nature* **432**, 353–360 (2004).
9. Vaziri, H. & Benchimol, S. Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.* **8**, 279–282 (1998).
10. Luo, J. *et al.* Negative control of p53 by Sir2α promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
11. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev.* **17**, 896–911 (2003).
12. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
13. Nishioka, K. *et al.* Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.* **16**, 479–489 (2002).
14. Kent, J. R. *et al.* During lytic infection herpes simplex virus type 1 is associated with histones bearing modifications that correlate with active transcription. *J. Virol.* **78**, 10178–10186 (2004).
15. Brown, M. A., Sims, R. J. III, Gottlieb, P. D. & Tucker, P. W. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol. Cancer* **5**, 26 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Barlev for the Set9 expression vector and D. Reinberg for the p53K372me1 antibody; T. Waibel for assistance in cloning Smyd2; S. Benchimol for the BJ and BJ-DNP53 cell lines; and members of the T.J. and S.L.B. laboratories for discussions. Research in the laboratory of T.J. is supported by the IMP through Boehringer Ingelheim and by grants from the European Union and the Austrian GEN-AU initiative, which is financed by the Austrian Ministry of Education, Science and Culture. Research support to S.L.B. was provided by a grant from the NIH. B.J.P. was supported by a Wistar Cancer Training Grant.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.B. (berger@wistar.org).

Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints

Jirina Bartkova^{1*}, Nousin Rezaei^{2*}, Michalis Liontos^{3*}, Panagiotis Karakaidos³, Dimitris Kletsas⁴, Natalia Issaeva⁵, Leandros-Vassilios F. Vassiliou³, Evangelos Kolettas⁶, Katerina Niforou³, Vassilis C. Zoumpourlis⁷, Munenori Takaoka⁸, Hiroshi Nakagawa⁸, Frederic Tort¹, Kasper Fugger¹, Fredrik Johansson⁵, Maxwell Sehested⁹, Claus L. Andersen¹⁰, Lars Dyrskjot¹⁰, Torben Ørntoft¹⁰, Jiri Lukas¹, Christos Kittas³, Thomas Hellday^{5,11}, Thanos D. Halazonetis^{2,12}, Jiri Bartek¹ & Vassilis G. Gorgoulis³

Recent studies have indicated the existence of tumorigenesis barriers that slow or inhibit the progression of preneoplastic lesions to neoplasia. One such barrier involves DNA replication stress, which leads to activation of the DNA damage checkpoint and thereby to apoptosis or cell cycle arrest^{1,2}, whereas a second barrier is mediated by oncogene-induced senescence^{3–6}. The relationship between these two barriers, if any, has not been elucidated. Here we show that oncogene-induced senescence is associated with signs of DNA replication stress, including prematurely terminated DNA replication forks and DNA double-strand breaks. Inhibiting the DNA double-strand break response kinase ataxia telangiectasia mutated (ATM) suppressed the induction of senescence and in a mouse model led to increased tumour size and invasiveness. Analysis of human precancerous lesions further indicated that DNA damage and senescence markers cosegregate closely. Thus, senescence in human preneoplastic lesions is a manifestation of oncogene-induced DNA replication stress and, together with apoptosis, provides a barrier to malignant progression.

There are several forms of senescence. Replicative senescence, the form induced by eroded telomeres, depends on activation of the DNA double-strand break (DSB) checkpoint kinases ATM and checkpoint kinase 2 (Chk2)^{7–9}. A second form of senescence induced by agents that cause DNA DSBs also depends on ATM and Chk2 (ref. 10). By contrast, oncogene-induced senescence, the form of senescence that is most likely to be associated with human precancerous lesions^{3,4}, has been linked to increased expression of the tumour suppressors p16^{INK4A} and ARF, rather than to activation of the DNA DSB checkpoint pathway^{11–13}.

To examine the possibility that, in addition to p16^{INK4A} and ARF, the DNA DSB checkpoint also contributes to oncogene-induced senescence, we studied human diploid cells expressing various oncogenes. The first oncogene we examined was *mos*, which was predicted to induce senescence, because, like *ras*, it activates the mitogen-activated protein kinase (MAPK) pathway^{14,15}. MRC5 diploid fibroblasts were infected with a retrovirus expressing *mos* or with a retrovirus without an insert, and transiently transduced cells were

selected. In these cells, overexpression of Mos protein was associated with phosphorylation of the extracellular signal-regulated kinase (ERK), indicating that the MAPK pathway was active, and also with p16^{INK4A} induction, activation of the DNA damage checkpoint and senescence-associated β -galactosidase (SA- β -gal) activity (Fig. 1a, b). To determine whether senescence was causally linked to activation of the DNA DSB checkpoint, we used short interfering RNA (siRNA) to deplete ATM, the key kinase responsible for DNA DSB signalling¹⁶. Senescence, as assayed by both SA- β -gal activity and incorporation of 5-bromodeoxyuridine (BrdU), was suppressed after ATM depletion (Fig. 1b–d). Similar results were obtained with BJ diploid fibroblasts and with a different pool of *atm*-specific siRNA oligonucleotides (see Supplementary Figs 1, 2).

The second oncogene we studied was *cdc6*. It encodes a DNA replication licensing factor that is expressed at high levels in non-small cell lung carcinoma (NSCLC) and other tumours^{17,18} and also in the MRC5 and BJ cells that overexpress Mos (Fig. 1a and Supplementary Fig. 1a). We generated MRC5, BJ and IMR90 cells expressing Cdc6 and observed activation of the DNA DSB checkpoint, as previously described¹⁹, and also induction of senescence (Supplementary Figs 3–5). The latter was suppressed by depleting the ATM kinase or its substrate p53 and also by caffeine, an ATM inhibitor²⁰ (Supplementary Figs 2–5).

The third oncogene we studied was *cyclin E*, which is frequently overexpressed in human precancerous and cancerous lesions²¹. We examined our previously described U2OS cells, in which cyclin E overexpression is induced by removal of tetracycline (Tet) from the tissue culture medium. Within 2 days, this leads to activation of the DNA DSB checkpoint¹. Similar to the findings described above for Mos and Cdc6, Cyclin E overexpression led to senescence, whose induction could be suppressed by a specific ATM inhibitor²² (Supplementary Fig. 6).

Our results indicate that the DNA DSB checkpoint mediates oncogene-induced senescence. However, previously, senescence mediated by *ras* was shown to involve p16^{INK4A} (refs 11–13). In our hands, overexpression of Mos and Cdc6 led to induction of p16^{INK4A} (Fig. 1a

¹Institute of Cancer Biology and Centre for Genotoxic Stress Research, Danish Cancer Society, DK-2100 Copenhagen, Denmark. ²The Wistar Institute, Philadelphia, Pennsylvania 19104-4268, USA. ³Department of Histology and Embryology, School of Medicine, University of Athens, GR-11527 Athens, Greece. ⁴Institute of Biology, Demokritos National Center for Scientific Research, GR-15310 Athens, Greece. ⁵Department of Genetics, Microbiology and Toxicology, Stockholm University, S-10691 Stockholm, Sweden. ⁶Department of Physiology, School of Medicine, University of Ioannina, GR-45110 Ioannina, Greece. ⁷Institute of Biological Research and Biotechnology, National Hellenic Research Foundation, GR-11635 Athens, Greece. ⁸Gastroenterology Division, Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-2144, USA. ⁹Department of Pathology, University Hospital, DK-2100 Copenhagen, Denmark. ¹⁰Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. ¹¹The Institute for Cancer Studies, University of Sheffield, Sheffield S10 2RX, UK. ¹²Department of Molecular Biology, University of Geneva, CH-1211 Geneva 4, Switzerland.

*These authors contributed equally to this work.

and Supplementary Figs 1a, 4a), but depletion of p16^{INK4A} by siRNA did not noticeably suppress senescence (Supplementary Fig. 7). Thus, the contribution of p16^{INK4A} to oncogene-induced senescence might vary from oncogene to oncogene, as it does for replicative senescence, which is mediated by the DNA DSB checkpoint, but also by a parallel pathway involving p16^{INK4A} (refs 7–9).

The mechanism by which the DNA DSB checkpoint is activated during oncogene-induced senescence could involve DNA replication stress. Consistent with this hypothesis, overexpression of Cdc6 led to senescence in MRC5 cells that express telomerase, indicating that the induction of senescence was not mediated by short telomeres (Supplementary Fig. 8). Furthermore, in BJ cells arrested in stage G1 of the cell cycle by serum starvation, overexpression of Mos or Cdc6 did not lead to activation of the DNA DSB checkpoint or senescence (Supplementary Fig. 9). To provide more evidence that

DNA replication stress is involved, we examined U2OS cells, which express cyclin E in an inducible manner, for single-stranded DNA, which is typically associated with stalled DNA replication forks, and for Chk1 phosphorylation on Ser 317 (ref. 23). Both markers scored positive (Supplementary Fig. 10a). Furthermore, the DNA damage foci in these cells, marked by H2AX phosphorylation, mapped to sites of DNA replication, as revealed by proliferating cell nuclear antigen (PCNA) staining (Supplementary Fig. 10b). Similarly, BJ cells overexpressing Mos and Cdc6 had replication protein A (RPA) foci, indicating the presence of single-stranded DNA (Supplementary Fig. 10c and data not shown).

We used DNA combing to investigate DNA replication stress further. U2OS cells expressing cyclin E in a Tet-repressible manner were pulse-labelled consecutively with the thymidine analogues IdU and CldU to distinguish ongoing DNA replication forks (a stretch of IdU label followed by CldU label) and newly-fired forks (CldU label only) from prematurely terminated forks (IdU label only). In cells that expressed normal levels of cyclin E (+Tet), most of the replication forks were ongoing or newly-fired, whereas, as early as one day after cyclin E induction (–Tet), most forks were prematurely terminated (Fig. 2a, b; Supplementary Fig. 11a, b). Premature termination of DNA replication can lead to fork collapse and formation of DNA DSBs. Indeed, overexpression of cyclin E led to the formation of DNA DSBs, but only in cells that were replicating their DNA (Fig. 2c). Similar results were obtained with BJ cells stably expressing cyclin E or Mos (Supplementary Fig. 11c–f).

The findings described here and also by d'Adda di Fagagna and co-workers²⁴ elucidate a previously undescribed pathway by which oncogenes induce senescence involving DNA replication stress and the DNA DSB checkpoint. We next sought to establish the relevance

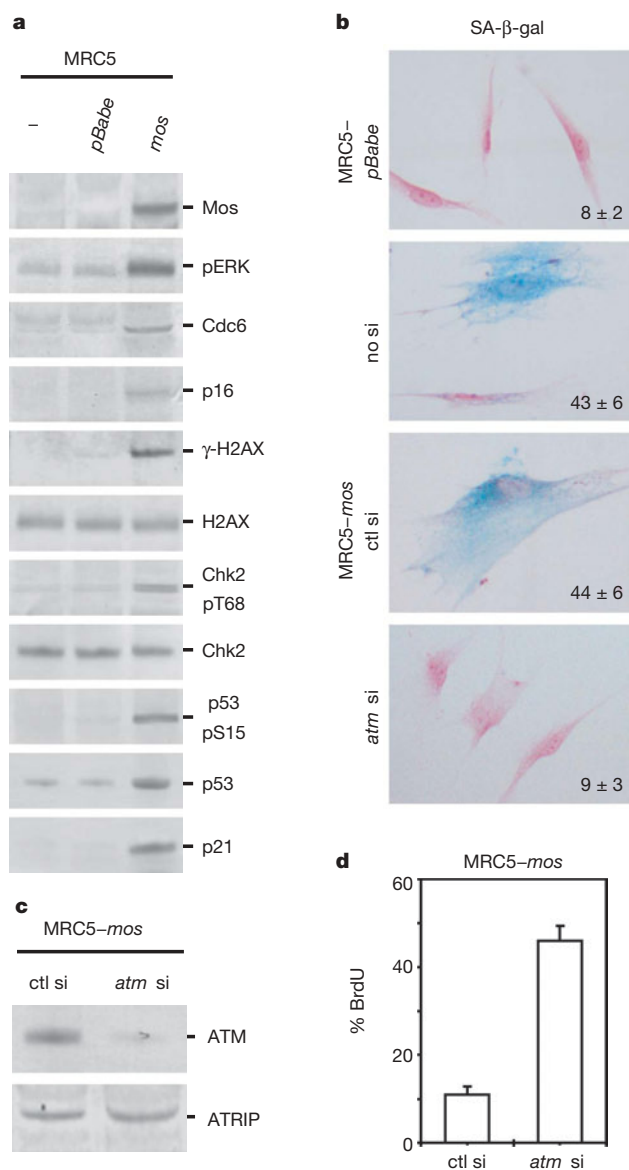


Figure 1 | Oncogene-induced senescence is dependent on the DNA DSB checkpoint. **a**, Immunoblot analysis of MRC5 diploid fibroblasts: non-infected (–), infected with empty vector (*pBabe*) or infected with a retroviral vector expressing *mos*. **b**, Images of SA-β-gal-stained MRC5 cells transfected with no siRNA (no si), control siRNA (ctl si) or siRNA specific for *atm* (*atm* si). Means and s.d. of the percentages of SA-β-gal-positive cells are indicated. **c**, Efficiency of ATM depletion. ATR-interacting protein (ATRIP) serves as a control. **d**, Means and s.d. of BrdU-positive staining for cells transfected with control or *atm*-specific siRNA.

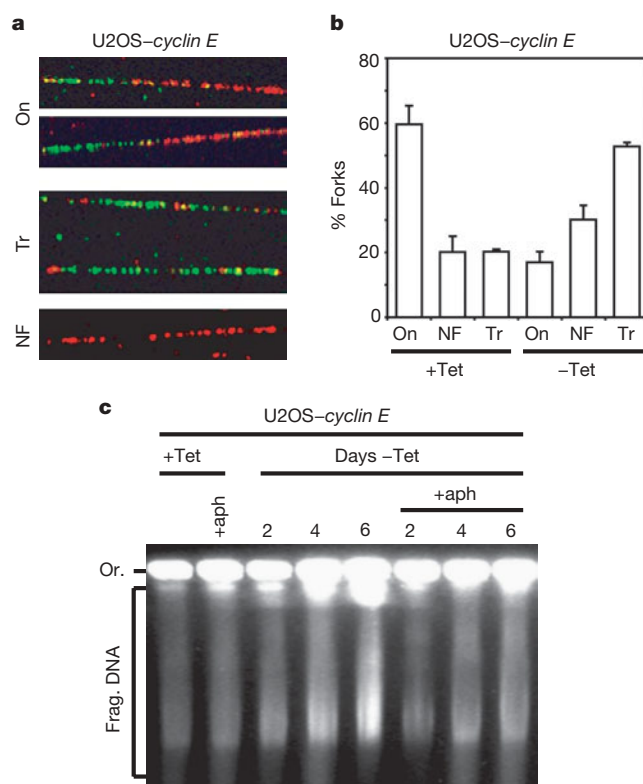


Figure 2 | DNA replication stress in cells undergoing oncogene-induced senescence. **a**, Representative DNA combing images of ongoing (On), terminated (Tr) and newly-fired (NF) replication forks. Cells were pulsed with IdU (green) and then with CldU (red). **b**, Mean and s.d. of the percentages of ongoing, newly-fired and terminated replication forks in cells with cyclin E repressed (+Tet) or induced (–Tet). **c**, Induction of cyclin E expression (–Tet) leads to DNA DSBs in cells undergoing DNA replication. aph, aphidicolin; Or., electrophoresis origin; Frag. DNA, fragmented DNA.

of this pathway in human cancerous and precancerous lesions. We first examined NSCLC, a tumour in which Cdc6 protein is overexpressed in about 50% of cases¹⁷. In 22 cases for which frozen sections were available from our previously described database^{2,17}, there was a remarkable correlation between Cdc6 overexpression, a potent DNA damage response, and SA- β -gal activity (Supplementary Fig. 12).

We subsequently examined preneoplastic lesions. Using frozen sections of colon adenomas, we first established a strong correlation between SA- β -gal staining and senescence markers suitable for immunohistochemistry of paraffin-embedded tissues, such as heterochromatin proteins 1 α and 1 γ (HP1 α and HP1 γ , respectively) and trimethylation of Lys 9 of histone H3 (refs 3–7, 12; Supplementary Fig. 13 and data not shown). This validation allowed us to study preneoplastic lesions for which only paraffin-embedded sections were available. In colon adenomas there was robust activation of the DNA DSB checkpoint that correlated very well with senescence (Fig. 3a and Supplementary Fig. 14). Progression to carcinoma was associated with decreased scores for both activation of the DNA DSB checkpoint and senescence. The decrease in senescence scores was

more pronounced, probably due to mutations in DNA damage checkpoint genes, such as *p53* (Fig. 3a). We found induction of p16^{INK4A} in both adenomas and carcinomas, but it did not correlate well with senescence (Fig. 3a). Furthermore, in those adenomas that showed heterogeneous staining, the presence of senescence correlated with activation of the DNA damage checkpoint, rather than with p16^{INK4A} induction (Fig. 3b). Analysis of precancerous and cancerous lesions of the urinary bladder led to similar findings (Supplementary Fig. 14). Finally, we also found an association between DNA damage response and senescence markers in benign mouse skin papillomas generated with the two step initiation/promotion protocol (Supplementary Fig. 15).

Senescence in human preneoplasia is thought to act as a barrier to tumorigenesis^{3–6}. The demonstration that oncogene-induced senescence is mediated by the DNA DSB checkpoint predicts that inhibition of the checkpoint would promote tumour progression. To test this hypothesis we established a tumour model using cells from the PDVC57 tumour line, injected subcutaneously in severe combined immunodeficient (SCID) mice and inhibiting ATM in these cells either by short hairpin (sh)RNA or with caffeine in the drinking water. PDVC57 cells have both copies of the endogenous *ras* gene mutated²⁵ and exhibit constitutive activation of the DNA DSB checkpoint and senescence. In tissue culture, ATM depletion by shRNA or inhibition by caffeine suppressed activation of the DNA DSB checkpoint and the induction of senescence, and led to increased invasiveness (Supplementary Fig. 16). Similarly, in tumours formed after injection of PDVC57 cells in SCID mice, depletion or inhibition of ATM suppressed activation of the DNA DSB checkpoint and senescence and led to larger and more invasive tumours (Fig. 4a–d and Supplementary Figs 17, 18).

Senescence was initially identified as a response of tissue culture cells to eroded telomeres or activated oncogenes¹³, but its relevance to human cancer was unclear until its recent demonstration in preneoplasia^{3–6}. The fact that senescence is more common in preneoplasia than in neoplasia indicated that senescence might serve as a barrier to oncogenesis. However, the molecular basis leading to induction of senescence in preneoplastic lesions was not defined and at least in melanocytic lesions did not involve telomere erosion or p16^{INK4A} induction³. Our results indicate that senescence, which, like apoptosis, serves as a tumorigenesis barrier in preneoplastic lesions, is induced by the DNA damage checkpoint in response to DNA replication stress (Fig. 4e).

METHODS

Manipulation of tissue culture cells. Human diploid IMR90 and MRC5 embryonic lung and BJ foreskin fibroblasts were infected at passages 8, 28 and 21, respectively, with a replication-incompetent retrovirus (*pBabe*) expressing human *mos* or *cdc6* or *cyclin E* and a hygromycin-resistance marker or with a control retrovirus expressing only the hygromycin-resistance marker. Stable clones (MRC5-*cdc6*, IMR90-*cdc6*, BJ-*cyclin E*) were selected with hygromycin (100 μ g ml⁻¹, Invitrogen) 48 h after infection over 3 weeks. Transiently transduced clones (MRC5-*mos*, BJ-*mos* and BJ-*cdc6*) were selected 24 h after infection for 2 days and examined 2 days later. U2OS cells expressing Tet-repressible cyclin E (ref. 1) were also examined. We used antibodies specific for the following proteins: Mos (clone P-19, Santa Cruz); pERK (ref. 25); Cdc6 (ref. 17); p16^{INK4A} (clone F-16, Santa Cruz); γ -H2AX, H2AX, Chk2 pT68, Chk2, p53, p21, ATRIP (ref. 2); ATM (GeneTex); Actin (clone AC-15, Abcam); RPA (clone 34-19, GeneTex); p53 pS15, cyclin E, Cdk7 (ref. 1); and Chk1 pS317 (ref. 26). Some cells were fixed with 1% paraformaldehyde, processed for SA- β -gal activity and counterstained with nuclear fast red²⁷. Cells were transfected with control siRNA (luciferase, Dharmacon) or with an siRNA pool specific for *atm* (ATM siGENOME, Dharmacon, sequences: 5'-CUA ACA AAC AGG UGA UAU AUU-3', 5'-AAA GGC CCA AGC UCC UCC UUU-3') after selection was completed (for the stable clones) or 2 days after selection was completed (for the transiently transduced clones). 72 h after siRNA transfection, the cells were assayed for SA- β -gal activity and BrdU incorporation²⁷.

DNA combing was performed as described previously²⁸ with slight modifications. Briefly, U2OS cells expressing Tet-repressible cyclin E were grown in the presence or absence of Tet for 1 day and then pulsed-labelled with 20 μ M IdU for

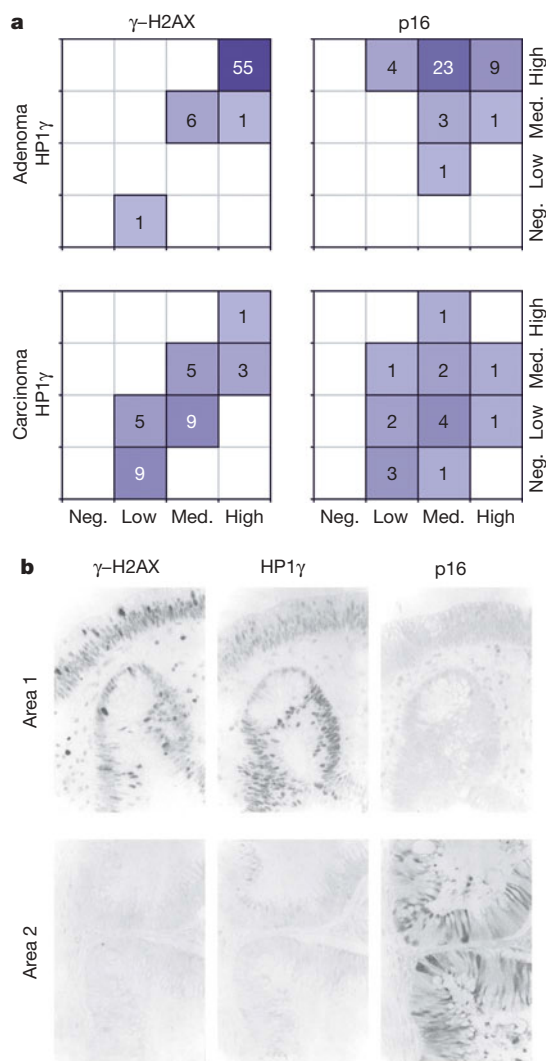


Figure 3 | Cosegregation of senescence and DNA damage response markers, but not p16^{INK4A}, in human precancerous and cancerous lesions. **a**, Correlation between HP1 γ expression, H2AX phosphorylation (γ -H2AX) and p16^{INK4A} (p16) expression in colon adenomas and carcinomas. For each marker, each lesion received a score of negative (Neg.), low, medium (Med.) or high, based on the percentage of cells scoring positive for that marker. **b**, Cosegregation of senescence and DNA damage response markers, but not p16^{INK4A} induction, in different areas of the same colon adenoma.

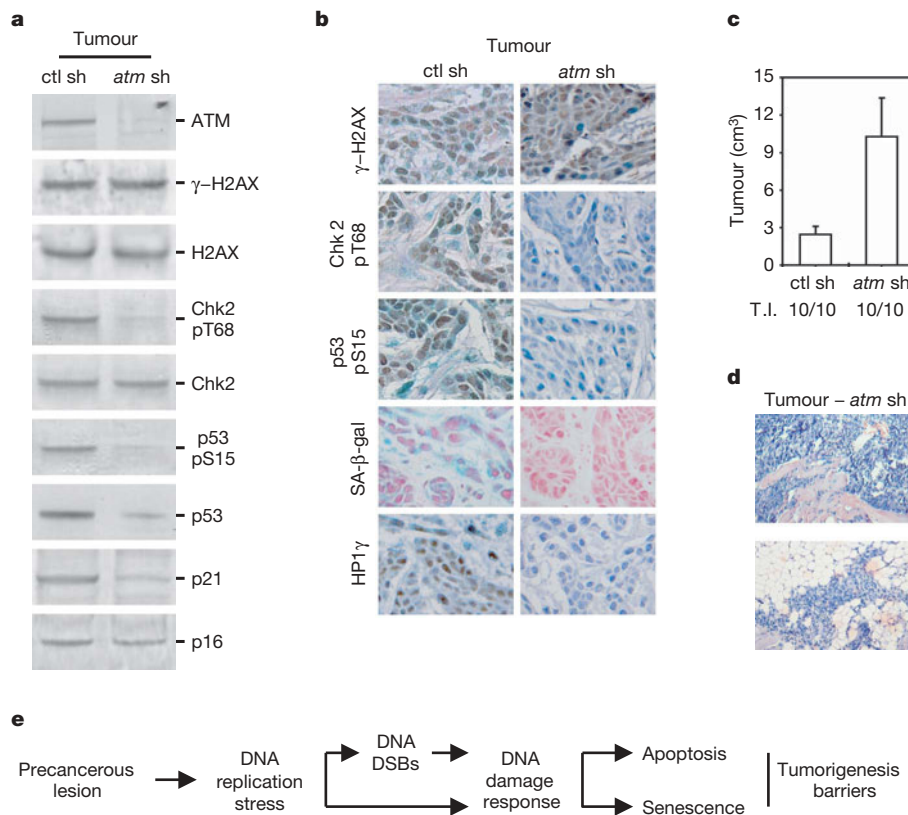


Figure 4 | Oncogene-induced senescence acts as a barrier to invasiveness and tumorigenesis. **a**, Immunoblot analysis of protein extracts prepared from tumours formed by PDVC57 cells stably expressing control shRNA (ctl sh) or *atm*-specific shRNA (*atm* sh). **b**, Immunohistochemistry and SA-β-gal analysis. **c**, Tumour volume (mean and s.d.) and tumour incidence (T.I.). T.I. indicates whether tumours formed in all injected sites. **d**, Histochemistry

showing invasion of PDVC57 cells expressing *atm*-specific shRNA (dark blue) in muscle tissue (pink; upper panel) and adipose tissue (clear; lower panel). **e**, Model for senescence and apoptosis as tumorigenesis barriers induced by activated DNA damage checkpoints in human precancerous lesions.

15 min, incubated in regular medium for 30 min, and then labelled with 20 μM CldU for 15 min. The cells were then harvested and lysed on glass slides in spread buffer. The DNA was denatured and stained with rat anti-BrdU/CldU (OBT0030F, Immunologicals Direct; 1:200) and mouse anti-IdU/BrdU (clone B44, Becton Dickinson; 1:50) primary antibodies.

For pulsed-field gel electrophoresis, U2OS cells expressing Tet-repressible cyclin E and BJ fibroblasts were treated with aphidicolin for 8 h and then allowed to grow for an additional 20 h (U2OS) or 40 h (BJ) in the absence or presence of aphidicolin. 1×10^6 cells were inserted into agarose plugs and subjected to pulsed-field gel electrophoresis²⁹.

Analysis of preneoplastic and neoplastic human lesions. Our databases of preneoplastic and neoplastic lesions have been described^{1,2,17}. Frozen sections of colon adenomas were from patients who had received no therapy before removal of the lesion. Immunohistochemistry was performed using optimally titrated antibodies specific for the following proteins: HP1α (15.19s2, Upstate); HP1γ (42s2, Upstate); histone H3 methylated on K9 (07-442, Upstate); p16^{INK4A} (16P04, NeoMarkers); Ki67, γ-H2AX, Chk2 pT68 (refs 1, 2). Each lesion was scored by an experienced pathologist as negative (no or only scattered stained cells, fewer than 2%), low (heterogeneous staining, with at least 20% of the section showing 2–10% positive cells), medium (at least 20% of the section showing 11–50% positive cells) or high (at least 20% of the section showing greater than 50% stained cells).

Mouse model for senescence as tumorigenesis barrier. PDVC57 cells were infected with a retrovirus expressing shRNA specific for *atm* (Open-Biosystems, shRNA sequence: 5'-TGC TGT TGA CAG TGA GCG CGC AGG GTC AGT CAA CAG ATT ATA GTG AAG CCA CAG ATG TAT AAT CTG TTG ACT GAC CCT GCA TGC CTA CTG CCT CGG A-3') or a control retrovirus and stably infected cells were selected over 3 weeks. To assay tumour progression, 10 SCID mice were injected subcutaneously at two different sites on their abdomen with 2×10^6 PDVC57 cells expressing shRNA specific for *atm* or with control cells³⁰. Two weeks later the mice were killed, the tumour volume was measured and the tumours were used to prepare protein extracts and for immunohistochemistry analysis and SA-β-gal staining. In a second experiment

the mice were killed 3 weeks after subcutaneous injection and tumour invasion was monitored by histochemistry.

Received 12 June; accepted 19 September 2006.

- Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
- Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
- Michaloglou, C. *et al.* BRAFE600-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
- Collado, M. *et al.* Tumour biology: senescence in premalignant tumours. *Nature* **436**, 642 (2005).
- Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
- Braig, M. *et al.* Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature* **436**, 660–665 (2005).
- Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
- d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
- Herbig, U., Jobling, W. A., Chen, B. P., Chen, D. J. & Sedivy, J. M. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol. Cell* **14**, 501–513 (2004).
- Di Leonardo, A., Linke, S. P., Clarkin, K. & Wahl, G. M. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev.* **8**, 2540–2551 (1994).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
- Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–716 (2003).
- Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* **432**, 307–315 (2004).
- Shibuya, E. K. & Ruderman, J. V. Mos induces the *in vitro* activation of mitogen-activated protein kinases in lysates of frog oocytes and mammalian somatic cells. *Mol. Biol. Cell* **4**, 781–790 (1993).

15. Lin, A. W. *et al.* Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. *Genes Dev.* **12**, 3008–3019 (1998).
16. Zgheib, O. *et al.* ATM signaling and 53BP1. *Radiother. Oncol.* **76**, 119–122 (2005).
17. Karakaidos, P. *et al.* Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant p53 on tumor growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. *Am. J. Pathol.* **165**, 1351–1365 (2004).
18. Williams, G. H. *et al.* Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proc. Natl Acad. Sci. USA* **95**, 14932–14937 (1998).
19. Vaziri, C. *et al.* A p53-dependent checkpoint pathway prevents rereplication. *Mol. Cell* **11**, 997–1008 (2003).
20. Sarkaria, J. N. *et al.* Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. *Cancer Res.* **59**, 4375–4382 (1999).
21. Hwang, H. C. & Clurman, B. E. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24**, 2776–2786 (2005).
22. Hickson, I. *et al.* Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res.* **64**, 9152–9159 (2004).
23. Lambert, S. & Carr, A. M. Checkpoint responses to replication fork barriers. *Biochimie* **87**, 591–602 (2005).
24. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage checkpoint response triggered by DNA hyper-replication. *Nature* (in the press).
25. Katsanakis, K. D., Gorgoulis, V., Papavassiliou, A. G. & Zoumpourlis, V. K. The progression in the mouse skin carcinogenesis model correlates with ERK1/2 signaling. *Mol. Med.* **8**, 624–637 (2002).
26. Raderschall, E., Golub, E. I. & Haaf, T. Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. *Proc. Natl Acad. Sci. USA* **96**, 1921–1926 (1999).
27. Ferbeyre, G. *et al.* PML is induced by oncogenic *ras* and promotes premature senescence. *Genes Dev.* **14**, 2015–2027 (2000).
28. Merrick, C. J., Jackson, D. & Diffley, J. F. Visualization of altered replication dynamics after DNA damage in human cells. *J. Biol. Chem.* **279**, 20067–20075 (2004).
29. Lundin, C. *et al.* Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable *in vivo* DNA double-strand breaks. *Nucleic Acids Res.* **33**, 3799–3811 (2005).
30. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Z. Lygerou and S. Taraviras for advice and reagents and M. Sideridou, N. Youroukos and M.-H. Lee for technical assistance. This work was supported by the Danish Cancer Society, the Danish National Research Foundation and the European Commission 'Active p53' and 'Mutant p53' Integrated Projects (J.B.); the National Cancer Institute, USA and the Swiss National Foundation (T.D.H.); the UICC (T.D.H. and V.G.G.); the Greek General Secretariat of Technology PENED program (V.G.G.); and the Swedish Cancer Society and Swedish Pain Relief Foundation (T.H.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.D.H. (Thanos.Halazonetis@molbio.unige.ch) or J.B. (jb@cancer.dk).

LETTERS

Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication

Raffaella Di Micco¹, Marzia Fumagalli¹, Angelo Cicalese², Sara Piccinin³, Patrizia Gasparini¹, Chiara Luise¹, Catherine Schurra⁴, Massimiliano Garre¹, Paolo Giovanni Nuciforo¹, Aaron Bensimon⁵, Roberta Maestro³, Pier Giuseppe Pelicci² & Fabrizio d'Adda di Fagnola¹

Early tumorigenesis is associated with the engagement of the DNA-damage checkpoint response (DDR)^{1,2}. Cell proliferation and transformation induced by oncogene activation are restrained by cellular senescence^{3–6}. It is unclear whether DDR activation and oncogene-induced senescence (OIS) are causally linked. Here we show that senescence, triggered by the expression of an activated oncogene (H-RasV12) in normal human cells, is a consequence of the activation of a robust DDR. Experimental inactivation of DDR abrogates OIS and promotes cell transformation. DDR and OIS are established after a hyper-replicative phase occurring immediately after oncogene expression. Senescent cells arrest with partly replicated DNA and with DNA replication origins having fired multiple times. *In vivo* DNA labelling and molecular DNA combing reveal that oncogene activation leads to augmented numbers of active replicons and to alterations in DNA replication fork progression. We also show that oncogene expression does not trigger a DDR in the absence of DNA replication. Last, we show that oncogene activation is associated with DDR activation in a mouse model *in vivo*. We propose that OIS results from the enforcement of a DDR triggered by oncogene-induced DNA hyper-replication.

Senescence is a barrier to cellular proliferation⁷. The molecular mechanisms that restrain cellular proliferation driven by activated oncogenes are not fully understood. DDR activation can arrest cell-cycle progression upon detection of DNA damage; we formed the hypothesis that DDR activation could contribute to OIS. Expression of oncogenic Ras (H-RasV12) induces OIS (ref. 8 and Supplementary Fig. 1). We therefore looked for cytological evidence of markers of activated DDR⁹ in normal human fibroblasts cultures that have undergone OIS by oncogenic Ras. Strikingly, we discovered that most cells undergoing OIS, but not control cells, have clearly detectable senescence-associated DNA-damage foci (SDF), containing the activated form of ataxia telangiectasia mutated (ATM; pS1981), NBS1, and the complex containing ATR (ATM- and Rad3-related), ATRIP (ATR-interacting protein) and RPA (replication protein A). SDF also include proteins phosphorylated by ATM or ATR (pS/TQ), activated DDR mediators such as 53BP1 and MDC1, together with phosphorylated cohesin SMC1 and histone H2AX (γ -H2AX; Fig. 1a, and Supplementary Fig. 2). OIS has been characterized by the formation of senescence-associated heterochromatic foci^{3–6,10}: the latter and SDF coexist in OIS cells but do not significantly colocalize (Supplementary Fig. 3). Further DDR analysis revealed a marked mobility shift of the DDR kinase CHK2 (checkpoint kinase 2) and phosphorylation on T68, an increase in CHK1 phosphorylation on S345, two activation events associated with DNA damage in potentially different cell cycle stages, and p53 phosphorylation on S15, together with an increase in the p21 and p16 cyclin-dependent kinase inhibitors, p53

and hypophosphorylated retinoblastoma (Rb) gene product (Fig. 1b). Similar observations were made in human fibroblasts transduced with human telomerase reverse transcriptase (hTERT), in mouse embryo fibroblasts (MEFs; data not shown) and in human mammary epithelial cells (HMECs; Supplementary Fig. 4). The detection of such a robust DDR correlates with the observed presence of DNA breaks in the majority of cells undergoing OIS as detected by TdT-mediated dUTP nick end labelling (TUNEL) and comet assays (Fig. 1c, d). We therefore conclude that OIS is associated with the presence of DNA damage and the full activation of the main DDR pathways.

To test whether the enforcement of a DDR has a causative function in the establishment of OIS, we expressed oncogenic Ras in DDR-deficient cells. Strikingly, whereas wild-type cells reduce their proliferation and senesce after Ras expression, cells in which CHK2 is stably inhibited by either of two distinct short hairpin RNAs (shCHK2-1 and shCHK2-2) continue to proliferate, despite Ras expression and activity (Fig. 2a–c, and Supplementary Figs 5–7). Similarly, ATM or p53 stable knockdown allows unrestrained proliferation (Fig. 2d–g, and Supplementary Figs 6 and 7). Notably, proliferating shCHK2 Ras cells accumulate DNA damage and retain activation of some DDR elements (Supplementary Fig. 8). Thus, efficient DDR activity is necessary to establish OIS, and its loss allows the proliferation of damaged cells.

Next we tested the effects of knockdown of DDR genes in cells in which OIS was fully established (Supplementary Fig. 9). Here, DDR inactivation results in an increase in the fraction of cells that restart DNA replication, as demonstrated by the incorporation of bromodeoxyuridine (BrdU) (Fig. 2h, and Supplementary Fig. 10), which is further enhanced by the additional inactivation of p16 (Fig. 2i). Furthermore, knockdown of the main DDR kinases allows efficient cellular proliferation, as indicated by increased cell numbers (Fig. 2j). Overall, these results show that the enforcement of a DDR is both causative and necessary in the establishment and in the maintenance of OIS.

We then tested whether DDR inactivation contributes to oncogene-induced cell transformation. We expressed oncogenic Ras in immortalized *CHK2*^{+/+} and *CHK2*^{-/-} MEFs. Loss of *CHK2* makes cells more susceptible to Ras-induced transformation, as revealed by altered cell morphology and loss of contact inhibition. Ras-expressing *CHK2*^{-/-} cells also show robust anchorage-independent growth in soft agar, whereas Ras-expressing *CHK2*^{+/+} cells form rare and small colonies (Fig. 3a, b). When injected subcutaneously in immunodeficient mice, Ras-expressing *CHK2*^{-/-} MEFs formed detectable tumours within two weeks (six out of six mice), whereas no tumours were observed in mice injected with Ras-expressing *CHK2*^{+/+} MEFs (zero out of six) at the same time point (Fig. 3c). DDR inactivation therefore promotes the transformation of mouse cells.

¹IFOM Foundation – FIRC Institute of Molecular Oncology Foundation, 20139 Milan, Italy. ²Department of Experimental Oncology, European Institute of Oncology, 20141 Milan, Italy. ³Experimental Oncology 1, Centro di Riferimento Oncologico CRO IRCCS, 33081 Aviano, Italy. ⁴Genome Stability Unit, 75724 Pasteur Institute, Paris, France. ⁵Genomic Vision, 75724 Paris, France.

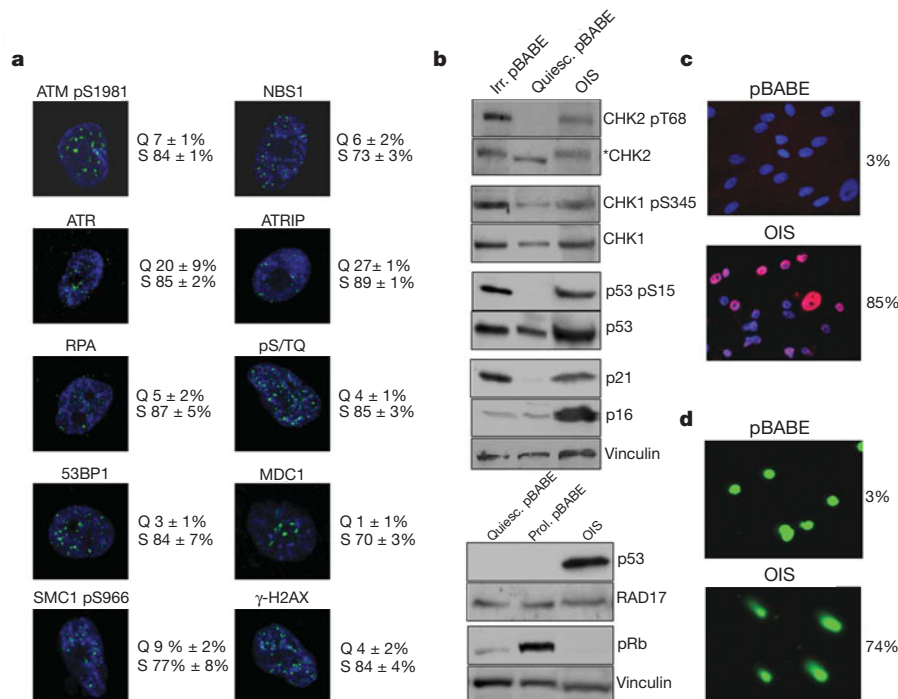


Figure 1 | DNA damage accumulation and DDR activation in OIS cells. **a**, SDF of the indicated DDR factors in OIS cells. Means $\pm$ s.d. indicate the fraction of SDF-positive cells in empty-vector (pBABE)-infected quiescent cells (Q) or senescent cells (S). Proliferating cells DDR is similar to quiescent cells. **b**, CHK2 pT68, CHK1 pS345, p53 pS15, p21, p16, p53 and

hypophosphorylated Rb accumulate in senescent cells; an asterisk highlights the altered mobility of CHK2. Rad17 and vinculin are loading controls. **c**, **d**, DNA breaks accumulate in OIS cells as demonstrated by TUNEL (**c**) and comet assay under denaturing conditions (**d**).

We also assayed the role of DDR in the transformation of human cells. Human fibroblasts are efficiently transformed by coexpression of Ras, E1A and MDM2 (mouse double minute 2)¹¹; we tested whether

DDR inactivation could functionally replace MDM2 expression. We discovered that either CHK2 or p53 inactivation cooperates with E1A and Ras expression to significantly increase both the number and the

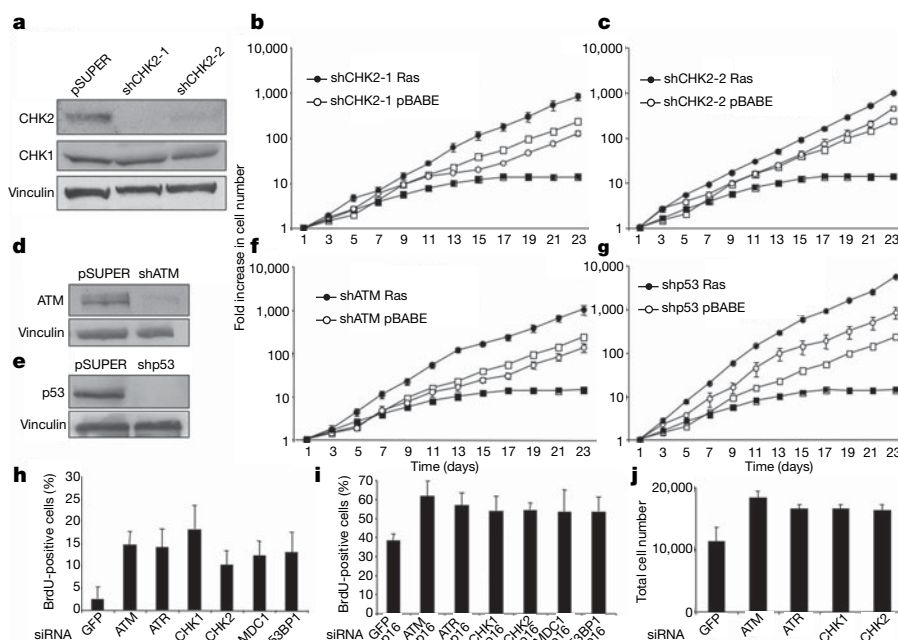


Figure 2 | DDR inactivation abolishes OIS. **a**, Stable downregulation of CHK2 by expression of two distinct short hairpin (sh) RNAs. CHK1 is unaffected. **b**, **c**, OIS is abolished in CHK2-downregulated cells as determined by growth curves. **d**, **e**, Stable downregulation of ATM or p53 respectively by shRNAs expression. **f**, **g**, Abolition of OIS in ATM or p53-downregulated cells as determined by growth curves. Open squares, pSUPER pBABE; filled squares, pSUPER Ras; open circles, shDDR gene

pBABE; filled circles, shDDR gene Ras. **h**, **i**, Transfection of siRNAs against DDR genes (**h**) and p16 (**i**) induces an increase in the fraction of BrdU-positive cells. **j**, Knockdown of DDR genes allows OIS cells to proliferate, as determined by increased cell numbers. Differences between siGFP (an siRNA against GFP) and siDDR individual genes are statistically significant: **h**, all $P < 0.003$; **i**, all $P < 0.05$; **j**, all $P < 0.04$. Error bars indicate s.d.; $n = 3$.

size of colonies in soft agar (Fig. 3d, e, and Supplementary Fig. 11). The stronger phenotype of p53-inactivated cells is consistent with redundant DDR (and other) pathways converging on p53. Thus, by the use of two defined systems, we conclude that DDR genes have tumour-suppressive functions.

The analysis of the events immediately after oncogenic Ras expression revealed that senescence is preceded by a strong proliferative burst (Fig. 4a); this is more evident in the absence of drug selection. A time-course study reveals that the end of the hyperproliferative phase is specifically associated in time with DDR engagement (Supplementary Fig. 12). Analysis by fluorescence-activated cell sorting (FACS) shows that a significant fraction of OIS cells arrest with a partly replicated DNA content (Supplementary Fig. 13), indicating the possible activation of an S-phase checkpoint. DNA re-replication—the firing of DNA replication origins more than once per cell cycle—has been shown to induce a DDR reminiscent of that observed in OIS cells¹². We therefore looked for evidence of DNA re-replication in OIS cells by fluorescence *in situ* hybridization (FISH) experiments with probes encompassing two distinct DNA replication origins (lamin B2 and β -globin¹³). Whereas we detected two signals per cell in control cells, a large fraction of OIS cells displayed more than two signals (Fig. 4b, and Supplementary Fig. 14). In addition, co-hybridization experiments consistently displayed a larger number of signals for the β -globin origin locus than for an unrelated centromeric alphoid probe on the same chromosome (Fig. 4b), which is consistent with the diploid DNA content profile observed by FACS.

We also analysed the pattern of whole-genome DNA replication *in vivo* in proliferating shCHK2 and shCHK2 Ras-expressing cells by molecular DNA combing¹⁴. DDR inactivation allows the proliferation of Ras-expressing cells without detectably affecting the numbers of altered FISH signals and diploidy (data not shown). We discovered that oncogene activation leads to an increased number of active replicons, as inferred by a smaller distance between origins (Fig. 4c, and Supplementary Fig. 15). This is consistent with the higher levels of CDC6 (cell division cycle 6 homologue), a positive DNA replication regulator, after Ras expression¹⁵ (Fig. 4d, and Supplementary Fig. 16); subsequent CDC6 downregulation is probably mediated by DDR¹⁶. DNA combing also revealed an increased asymmetry in the progression between right and left forks emanating from the same origin in Ras-expressing cells; such discontinuous fork advancement may be the consequence of increased fork instability or extensive fork pausing (Fig. 4c and Supplementary Fig. 15).

Fragile sites are slow-replicating genomic regions that are prone to engage the ATR pathway¹⁷. Single-nucleotide polymorphism analysis at the three most common fragile sites in control and OIS cells reveals that loss of heterozygosity (LOH, a measure of genome instability) is more likely to occur here than in the rest of the genome in OIS cells (Fig. 4e). Overall, all the available data indicate that oncogene expression leads to an aberrant increase in DNA replication events, genome instability and DDR activation.

Our results support the idea that DNA replication is crucial for DDR induction by oncogenes. To test this hypothesis, we induced oncogene

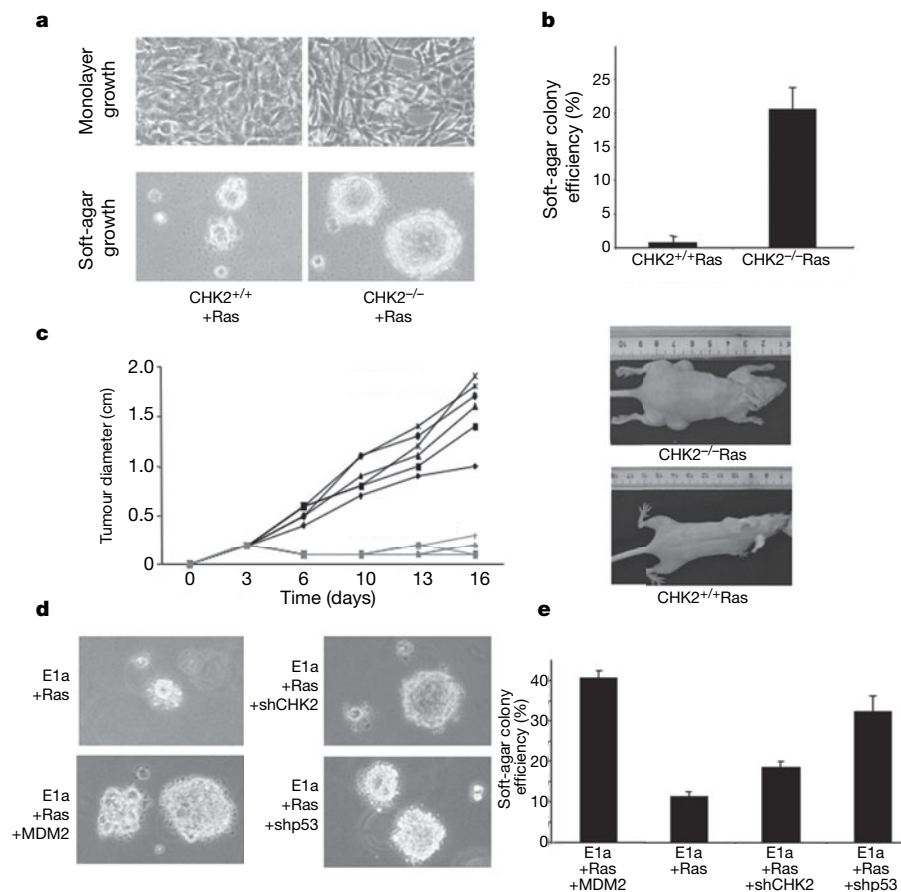


Figure 3 | DDR inactivation promotes cell transformation. **a**, Oncogenic Ras expression induces a transformed phenotype in $CHK2^{-/-}$ mouse cells but not in $CHK2^{+/+}$ cells, as shown by spindle-like morphology and loss of contact inhibition (top) and anchorage-independent growth in soft agar (bottom). **b**, Quantification of colony formation efficiency (percentage of colonies per 100 plated cells; $P < 10^{-10}$, $n = 19$). **c**, *In vivo* tumour growth rates of Ras expressing MEFs injected subcutaneously in immunocompromised mice are enhanced by loss of $CHK2$ ($n = 6$). Grey,

$CHK2^{+/+}$ Ras; black, $CHK2^{-/-}$ Ras. Error bars indicate s.e.m. Photographs show two representative immunocompromised mice injected with the indicated cells. **d**, shCHK2 expression in human cells cooperates with E1a and oncogenic Ras to increase the number and the size of cell colonies in soft agar. **e**, Quantification of colony formation efficiency. Differences between E1a + Ras and E1a + Ras + shCHK2 are statistically significant: $P < 0.01$. Error bars indicate s.e.m.; $n = 5$.

expression in cells unable to undergo DNA replication, either because they were contact-inhibited or because they had been released from G0 in the presence of aphidicolin, a DNA polymerase inhibitor, and in control replicating cells. We discovered that oncogene expression in cells unable to undergo DNA replication does not trigger a detectable DDR (Fig. 4f). This is despite comparable levels of Ras expression and activation (Supplementary Fig. 17). Therefore, oncogene-induced DDR and senescence are dependent on DNA replication.

We extended our observations *in vivo* by the use of a mouse model (*Mus musculus*) of chemically induced skin carcinogenesis associated with H-Ras-activating mutations. Treatment with 7,12-dimethylbenz (a)anthracene (DMBA) plus 12-*O*-tetradecanoylphorbol-13-acetate (TPA) induces hyperproliferative papillomatous skin lesions¹⁸ and accumulation of OIS markers⁶. We observed γ -H2AX-positive cells

among skin keratinocytes in the basal and suprabasal layers of all early benign skin papillomas studied, but not in the treated skin immediately adjacent to the lesions and in mock-treated mice (Fig. 4g, and Supplementary Fig. 18). Therefore, cellular hyperproliferation associated with activating mutations of Ras triggers a DDR also *in vivo*.

Thus, our data indicate that OIS results from the enforcement of a robust DDR. The observation that DDR functions in both senescence induced by telomere shortening^{19,20} and oncogene activation indicates that activation of some DDR elements may be a general feature of cellular senescence. We propose a model (Fig. 4h) in which oncogene expression leads to hyperproliferation and DNA hyper-replication. This causes an accumulation of DNA damage and sparks the activation of an S-phase-specific DDR. In DDR-proficient cells, this results in cellular senescence. In the absence of effective DDR

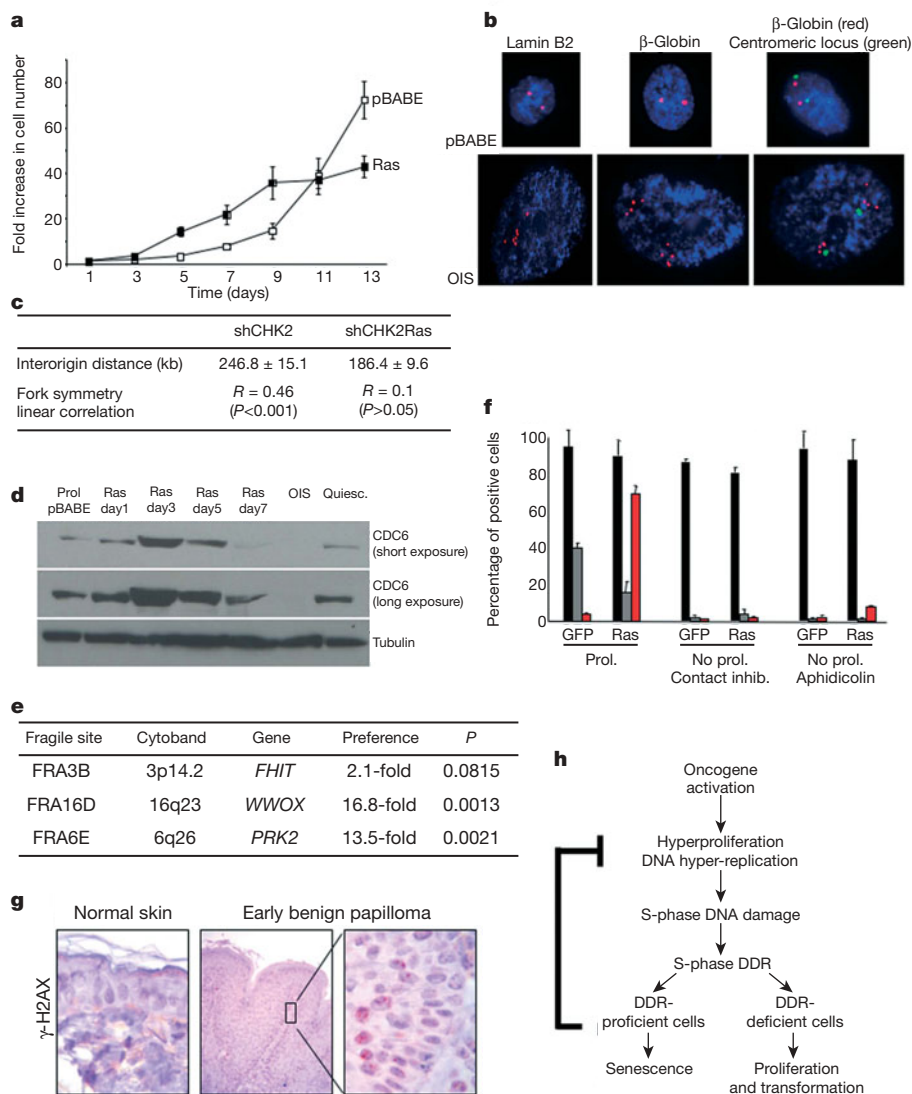


Figure 4 | Oncogene expression leads to alterations of DNA replication.

a, Growth curve shows faster cell proliferation immediately after oncogene expression than control cells. Error bars indicate s.d. ($n = 3$). **b**, OIS cells display multiple FISH signals for DNA replication origins. Signals generated by an unrelated centromeric alphoid probe (PRB11) on the same chromosome 11 are fewer than β -globin origins in OIS cells. **c**, Ras expression increases the number of active replicons but leads to an asymmetric fork progression, as detected by DNA combing. **d**, Immunoblotting shows that oncogene expression leads to a transient upregulation of CDC6. **e**, Single-nucleotide

polymorphism analysis reveals significant loss-of-heterozygosity (LOH) preference at two of the three most common fragile sites in OIS cells. **f**, DDR activation by oncogenic Ras, as detected by pS/TQ immunostaining, is strongly reduced in contact-inhibited or aphidicolin-blocked cells; 100 cells were screened in each sample. Error bars indicate s.d. ($n = 3$). Black bars, transduction efficiency; grey bars, BrdU incorporation; red bars, pS/TQ; inhib., inhibited; prol., proliferation. **g**, Treatment with DMBA/TPA triggers DDR *in vivo* in early benign skin papillomas. **h**, Proposed model of the events after oncogene expression.

mechanisms, oncogene expression fuels unrestrained proliferation and contributes to cell transformation.

OIS cells have a consistent set of markers of an activated DNA replication checkpoint such as a high fraction of cells arrested in S phase, activation of the ATR pathway and preferential loss of heterozygosity at fragile sites. Oncogenic Ras expression leads to overexpression of CDC6, to increased numbers of active replicons and to OIS cells displaying multiple DNA-replication origin signals. In agreement with our results, oncogenic Ras has been shown to induce genome instability after a single round of DNA replication²¹, CDC6 and MCM (mini-chromosome maintenance) DNA replication factor proteins are overexpressed in some tumours^{22,23} and CDC6 overexpression is sufficient to induce DDR activation and senescence^{12,24}, although it may have additional oncogenic functions²⁵.

DDR activation has a causative function in OIS, as shown by the observation that individual inactivation of several DDR factors can both prevent OIS and allow escape from it. In addition, CHK2 inactivation promotes cellular transformation, in agreement with the increased susceptibility of *CHK2*^{-/-} mice to tumours induced by DMBA/TPA²⁶. Our results, together with those reported by ref. 24, strongly suggest a tumour-suppressive role of DDR signalling in OIS. Because the activation of several additional oncogenes such as *Myc*²⁷, cyclin E, *Cdc25A*, *E2F1* and *Mos*^{1,24} or the inactivation of *PTEN* (phosphatase and tensin homologue deleted on chromosome 10) engages the DDR machinery²⁸, it is tempting to propose our conclusions as a more general paradigm of oncogenesis.

METHODS

Detailed experimental methods and a list of associated protocols are described in Supplementary Information.

Cell culture. BJ (ATCC) and BJ hTERT (Clontech) normal human fibroblasts and HMECs (Clonetics) cells were grown under standard tissue-culture conditions. pRETROSUPER plasmids were from the SUPER RNAi library. The β -galactosidase assay was performed as described²⁹. Comet and TUNEL assays were performed with CometAssay (Trevigen) and *in situ* Cell Death Detection (Roche) kits, respectively, in accordance with the manufacturers' instructions; 50–100 cells were screened in each sample.

Transformation assays. Cells were transduced with retroviral vectors and a vector encoding enhanced green fluorescent protein to monitor infection efficiency and analysed for anchorage-independent growth after two weeks in semisolid medium. For tumorigenicity assays, 5×10^6 cells were injected subcutaneously into the flanks of the immunocompromised mice.

Immunofluorescence microscopy. Cells were fixed and probed as described¹⁹. ATR, ATRIP and RPA staining was preceded by *in situ* cell fractionation³⁰. Confocal sections were obtained with a Leica TCS SP2 AOBs confocal laser microscope by sequential scanning. Comparative immunofluorescence analyses were performed in parallel with identical acquisition parameters; 100–300 cells were screened for each antigen.

siRNA. Sequences of RNA oligonucleotides are described in Supplementary Information. Oligofection efficiency was monitored by the use of a fluorescein isothiocyanate-labelled siRNA and found to be about 70%. Bromodeoxyuridine incorporation and cell proliferation were analysed three days after transfection.

FISH. FISH was performed as described¹³. Probes were bacterial artificial chromosome (BAC) clone no. RP11-211I3 (containing the lamin B2 replication origin) and BAC clone no. RP11-645I8 (containing the β -globin replication origin) from the Children's Oakland Research Institute. For each point, 200–300 cells were scored.

Statistical analysis. Results shown are means $\pm$ s.d or means $\pm$ s.e.m. as indicated. *P* was calculated by Student's two-tailed *t*-test.

Received 12 July; accepted 10 October 2006.

1. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
2. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
3. Michaloglou, C. *et al.* BRAF600-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
4. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
5. Braig, M. *et al.* Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature* **436**, 660–665 (2005).
6. Collado, M. *et al.* Tumour biology: senescence in premalignant tumours. *Nature* **436**, 642 (2005).

7. MATHON, N. F. & LLOYD, A. C. Cell senescence and cancer. *Nature Rev. Cancer* **1**, 203–213 (2001).
8. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic Ras provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
9. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
10. Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–716 (2003).
11. Seger, Y. R. *et al.* Transformation of normal human cells in the absence of telomerase activation. *Cancer Cell* **2**, 401–413 (2002).
12. Vaziri, C. *et al.* A p53-dependent checkpoint pathway prevents rereplication. *Mol. Cell* **11**, 997–1008 (2003).
13. Melixietian, M. *et al.* Loss of Geminin induces rereplication in the presence of functional p53. *J. Cell Biol.* **165**, 473–482 (2004).
14. Lebofsky, R. & Bensimon, A. Single DNA molecule analysis: applications of molecular combing. *Brief. Funct. Genomic. Proteomic.* **1**, 385–396 (2003).
15. Takeda, D. Y., Shibata, Y., Parvin, J. D. & Dutta, A. Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells. *Genes Dev.* **19**, 2827–2836 (2005).
16. Duursma, A. & Agami, R. p53-dependent regulation of Cdc6 protein stability controls cellular proliferation. *Mol. Cell Biol.* **25**, 6937–6947 (2005).
17. Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).
18. Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322**, 78–80 (1986).
19. d'Adda di Fagnano, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
20. Herbig, U., Jobling, W. A., Chen, B. P., Chen, D. J. & Sedivy, J. M. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21^{CIP1}, but not p16^{INK4a}. *Mol. Cell* **14**, 501–513 (2004).
21. Denko, N. C., Giaccia, A. J., Stringer, J. R. & Stambrook, P. J. The human Ha-ras oncogene induces genomic instability in murine fibroblasts within one cell cycle. *Proc. Natl Acad. Sci. USA* **91**, 5124–5128 (1994).
22. Williams, G. H. *et al.* Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proc. Natl Acad. Sci. USA* **95**, 14932–14937 (1998).
23. Karakaidos, P. *et al.* Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: Synergistic effect with mutant p53 on tumor growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. *Am. J. Pathol.* **165**, 1351–1365 (2004).
24. Bartkova, J. *et al.* Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* doi:10.1038/nature05268 (this issue).
25. Gonzalez, S. *et al.* Oncogenic activity of Cdc6 through repression of the *INK4A/ARF* locus. *Nature* **440**, 702–706 (2006).
26. Hirao, A. *et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Mol. Cell Biol.* **22**, 6521–6532 (2002).
27. Vafa, O. *et al.* c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability. *Mol. Cell* **9**, 1031–1044 (2002).
28. Puc, J. *et al.* Lack of PTEN sequesters CHK1 and initiates genetic instability. *Cancer Cell* **7**, 193–204 (2005).
29. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
30. Mirzoeva, O. K. & Petrini, J. H. J. DNA damage-dependent nuclear dynamics of the Mre11 complex. *Mol. Cell Biol.* **21**, 281–288 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Transidico for confocal analysis; IFOM Cell Biology, DNA Sequencing, Imaging and Microarrays Units for support; J. Campisi, T. D. Halazonetis, K. Helin, S. P. Jackson, Y. Shiloh, T. Mak, C. Marine, S. Pece and A. Tocchetti for sharing reagents; and M. Alcalay, B. Amati, A. Ballabeni, M. Barbacid, M. Foiani and H. Muller for discussions. F.d.A.d.F., R.M. and P.G.P. are supported by AIRC (Associazione Italiana per la Ricerca sul Cancro). R.D.M. and M.F. are SEMM (European School of Molecular Medicine) students.

Author Contributions M.F. generated Fig. 4f and Supplementary Figs 4c, 7 and 17; A.C. generated Supplementary Figs 51a, b and 4a, b and helped with initial retroviral transduction experiments; S.P. and R.M. generated and analysed data in Fig. 3 and Supplementary Fig. 11; P.G. generated Fig. 4b and Supplementary Fig. 14; C.L. and P.G.N. generated and analysed data in Fig. 4g; C.S. and A.B. generated and analysed data in Fig. 4c and Supplementary Fig. 15; M.G. helped with confocal analysis; P.G.P. discussed the results; R.D.M. generated all remaining figures and contributed to experimental design; F.d.A.d.F. generated Fig. 4e, planned the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.d.A.d.F. (fabrizio.dadda@ifom-ieo-campus.it).

Variability and memory of protein levels in human cells

Alex Sigal^{1*}, Ron Milo^{1*†}, Ariel Cohen^{1*}, Naama Geva-Zatorsky¹, Yael Klein¹, Yuval Liron¹, Nitzan Rosenfeld¹, Tamar Danon¹, Natalie Perzov¹ & Uri Alon¹

Protein expression is a stochastic process that leads to phenotypic variation among cells^{1–6}. The cell–cell distribution of protein levels in microorganisms has been well characterized^{7–23} but little is known about such variability in human cells. Here, we studied the variability of protein levels in human cells, as well as the temporal dynamics of this variability, and addressed whether cells with higher than average protein levels eventually have lower than average levels, and if so, over what timescale does this mixing occur. We measured fluctuations over time in the levels of 20 endogenous proteins in living human cells, tagged by the gene for yellow fluorescent protein at their chromosomal loci²⁴. We found variability with a standard deviation that ranged, for different proteins, from about 15% to 30% of the mean. Mixing between high and low levels occurred for all proteins, but the mixing time was longer than two cell generations (more than 40 h) for many proteins. We also tagged pairs of proteins with two colours, and

found that the levels of proteins in the same biological pathway were far more correlated than those of proteins in different pathways. The persistent memory for protein levels that we found might underlie individuality in cell behaviour and could set a timescale needed for signals to affect fully every member of a cell population.

We asked whether and on what timescale do protein levels mix in individual human cells. Mixing, in the present context, occurs when a cell lineage, given enough time, reaches the different states found in a snapshot of a cell population. Irreversible differences between cells, such as expected from genetic changes, would lead to cells that retain their protein state indefinitely without mixing (Fig. 1a, left). Mixing behaviour (Fig. 1a, middle and right) can be characterized by a typical timescale τ_m , called the mixing time, after which cells lose the memory of their previous protein state. Recent studies using synthetic protein expression systems in bacteria indicate that cells with

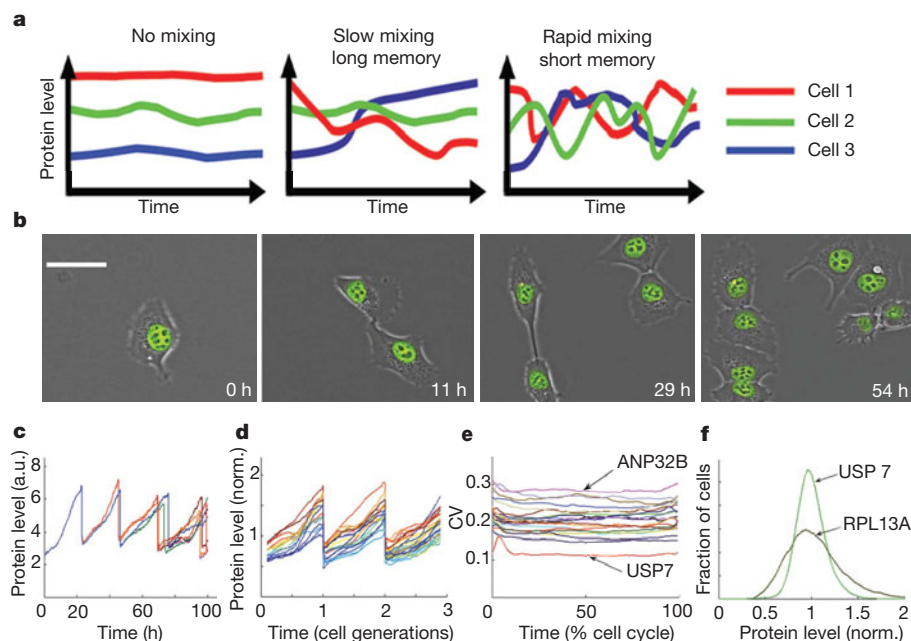


Figure 1 | Protein level dynamics in individual cells. **a**, Possible mixing dynamics of a protein in a population of cells. **b**, Cells expressing YFP CD-tagged topoisomerase 1 (TOP1) at different time-points. All cells are the progeny of the cell shown in the first frame. Scale bar is 25 μ m. **c**, TOP1 levels as a function of time in a cell lineage. Sharp decreases in protein levels indicate cell-division events and lines originating after each division are the

TOP1 levels of daughter cells. **d**, TOP1 protein dynamics synchronized to a cell-cycle time-base using cell division events. Protein levels were normalized by the average protein level. **e**, The coefficient of variation ($CV = s.d./mean$) across the cell cycle of 20 different proteins. **f**, Distribution of protein levels measured by flow cytometry for cells at a similar cell-cycle stage, for USP7 ($CV = 0.16$) and RPL13A ($CV = 0.28$).

¹Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, 76100 Israel. [†]Present address: Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

high production tend to maintain this for about a cell generation¹², yielding mixing times on the order of 1–2 h (ref. 13). However, little quantitative information exists about variability in protein levels among human cells, and specifically among human cancer cells, where acquired genetic differences might add to non-genetic sources of noise.

To examine the extent and persistence of variability in the levels of different proteins, we used a fluorescent labelling strategy known as CD-tagging, which tags proteins at their endogenous chromosomal locations^{24–26}. A yellow fluorescent protein (YFP) sequence was retrovirally inserted into the genome of human H1299 lung carcinoma cells. The YFP sequence contained splice acceptor and donor flanking sequences, which enabled the tag to be spliced into the mRNA of a gene as a new exon, resulting in the expression of a fusion protein^{24,25}. The tagged gene was therefore expressed under its native promoter on the chromosome. CD-tagging has been shown to preserve the localization and function of most of the tagged proteins^{24–27}.

The dynamics of YFP-tagged proteins in individual cells were followed by time-lapse fluorescence microscopy under controlled temperature, humidity and CO₂ conditions (Fig. 1b, see also Supplementary Movie 1 (transmitted light) and 2 (fluorescence) of cells shown in Fig. 1b). Time-lapse movies were taken over 50–100 h, at a resolution of 1 frame per 10 min.

We quantified protein levels in individual cells over multiple cell cycles (Fig. 1c). We used proteins that were localized in the nucleus to facilitate accurate automatic image analysis of the movies (see Methods). The twenty proteins studied were in a variety of pathways, including RNA splicing, transcriptional regulation, apoptosis and DNA replication (see Supplementary Information). All proteins studied were correctly localized.

To quantify the inherent variability between cells, we synchronized cells *in silico* by aligning protein dynamics between division events²⁴. *In silico* synchronization enabled us to examine variability in cells in which an equal fraction of the cell cycle has elapsed, eliminating differences in cell-cycle stage^{7,18} as a source of variability (Fig. 1d). We found that the average coefficient of variation of the total fluorescence (CV = standard deviation/mean) depended on the protein

measured and ranged from 0.12 to 0.28 (Fig. 1e and Supplementary Information). Protein level distributions were single-peaked and skewed to the right (Fig. 1f). The CV values were nearly constant throughout the cell cycle and were comparable to those found in bacteria and yeast (0.1–1.0)^{7,11,14,20}. The ratio between the fluorescence of cells at the 90th percentile to cells at the 10th percentile ranged from about 1.3 to 2.1 (Supplementary Information). The density of cells in our experiments, as measured by the number of cell–cell contacts, did not significantly contribute to the variability in the levels of the proteins examined (see Supplementary Information).

We next studied the mixing dynamics of these proteins. For each protein, we followed a clonal population of cells over several generations using time-lapse microscopy, and synchronized cells to the cell-cycle time base (Fig. 2a, b). For some proteins, such as the ubiquitin-specific protease USP7, mixing occurred in about one cell cycle (~20 h; Fig. 2c). By contrast, other proteins, such as the transcription-regulating factor HMGA2, showed slower mixing, on the time-scale of several cell generations (Fig. 2d). Cell division events did not seem to have a strong randomizing effect on protein levels (Fig. 2c, d), in contrast to observations on mRNA levels in bacteria²⁸.

To compute the mixing time, we calculated the auto-correlation function $A(\tau)$ of the protein levels (Fig. 2e, f and see Methods). The mixing time τ_m was defined as the time when $A(\tau)$ decayed to one half. We found that τ_m ranged from about 0.8 generations for CIRBP, a protein involved in the cold-shock response, to about 2.5 generations for HMGA2 (see Supplementary Information). Furthermore, there was a substantial correlation ($R^2 \approx 0.6$) between the cell–cell variability in protein levels (CV) and the mixing time (Fig. 3). This correlation between mixing time and variability might pose constraints on possible mechanisms for protein noise (see Supplementary Information).

We also investigated the correlations between different proteins in the same cell. For this, we used CD-tagging to generate clones in which two proteins are tagged, one with yellow (YFP) and the other with a red (*mCherry*²⁹) fluorescent tag. The levels of randomly chosen pairs of proteins from different pathways were nearly uncorrelated

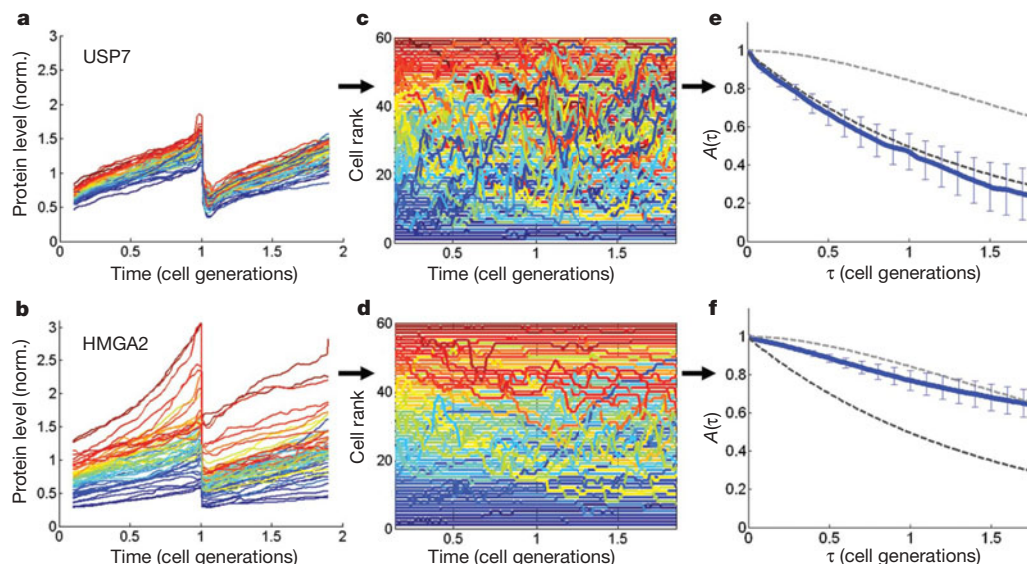


Figure 2 | Mixing-times of different proteins. **a, b**, Dynamics of cells expressing YFP CD-tagged USP7 and HMGA2 across two cell generations. Each line represents the protein level of one cell, normalized to the population mean. **c, d**, Ranks of 60 randomly selected cells for USP7 and HMGA2 through two cell generations, where 1 is the lowest and 60 is the highest protein level in the population at the given time-point. Lines are colour-coded for rank at the start of the first cell cycle, with red being high relative to the mean and blue low. **e, f**, The auto-correlations of protein level

ranks, calculated on the basis of at least 150 cells for each protein. The lower dashed line is the theoretical auto-correlation for stable proteins with noisy production, and dilution by cell growth ($A(\tau) = e^{-\alpha\tau}$, where α is the growth rate). The upper dashed line is the theoretical auto-correlation for proteins whose production rate depends on an upstream stable protein with noisy production, $A(\tau) = e^{-\alpha\tau} \cdot (1 + \alpha\tau)$. For theoretical models, see Supplementary Information. Bars denote standard errors (bootstrap method).

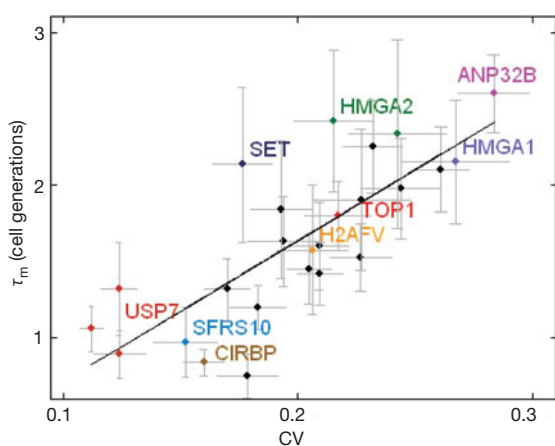


Figure 3 | Mixing time correlates with variability. Mixing time (τ_m) versus CV (s.d./mean) for the 20 proteins described in Supplementary Table 1. The best fit line is shown ($R^2 = 0.62$). Independently derived clones of USP7 and HMGA2 are in red and green. Bars denote standard errors (bootstrap method).

(Fig. 4a, d; see Supplementary Information for functions of proteins studied). The levels of proteins from the same system (ribosome) or levels of proteins produced from different alleles of the same gene (ribosomal protein RPL5) were substantially correlated (Fig. 4b–d). These findings indicate that the main source of noise might be variability in upstream regulatory components specific to each system. Such noise accounted for about 80% of the variance ($\eta_{\text{ext}}^2/\eta_{\text{tot}}^2$) in levels of RPL5, consistent with measurements of high abundance proteins in microorganisms^{7–9,14}.

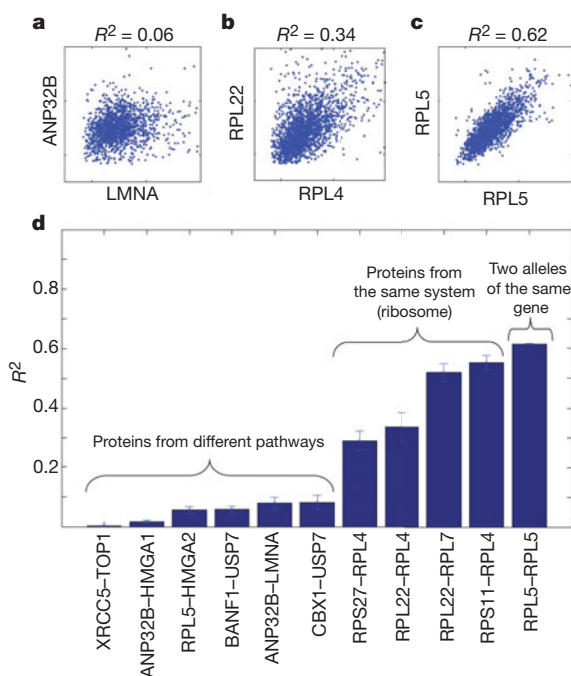


Figure 4 | Correlations between pairs of proteins CD-tagged with two fluorescent colours (YFP and mCherry) in the same cells. Shown is mean nuclear fluorescence of each protein normalized to its population average. **a**, Two proteins with different biological functions: LMNA (component of nuclear membrane) and ANP32B (involved in apoptosis). **b**, Two proteins in the same biological system (two ribosomal proteins, RPL4 and RPL22). **c**, Products of different alleles of the same gene, RPL5. The extrinsic and intrinsic noise components⁷ of RPL5 are $\eta_{\text{ext}} = 0.28$ and $\eta_{\text{int}} = 0.14$, and $\eta_{\text{tot}} = 0.31$. **d**, All protein pairs measured and their correlations. Bars denote mean $R^2 \pm$ s.e. from three experiments performed on different days.

We compared our finding of long mixing times to simple theoretical models in which proteins are produced and degraded stochastically and diluted out by cell growth (see Supplementary Information). Theoretical analysis leads to the following qualitative features: (1) simple production–degradation models can give mixing times that are at most one cell generation; (2) active protein degradation can reduce the mixing time; and (3) mixing times longer than a cell generation can occur as a result of propagation of noise from upstream components, as suggested by the above correlations observed in proteins from the same pathway, or due to feedback loops.

We found that cell–cell variability in human protein levels can have a long mixing time: proteins whose levels are higher or lower than average can remain so for several days. Such long-memory noise can act, through the cells' regulatory circuitry^{3,11,21}, to generate variability in cellular responses and phenotypes. In particular, proteins from co-regulated systems such as the ribosome seem to fluctuate in a much more correlated fashion than proteins from different pathways. This long-lasting, concerted noise might result, for example, in 'outlier' cells that have different behaviour from the bulk of the population. The observed long memory indicates that several cell cycles might be needed for some signals and drugs to affect fully populations of human cancer cells.

METHODS

YFP CD-tagging of endogenous proteins. A library of CD-tagged proteins in the H1299 non-small cell lung carcinoma cell line was constructed as described²⁴. Briefly, *EYFP*, flanked with splice acceptor and donor sequences, was integrated into the genome using pBabeAE retroviral vectors, each containing *EYFP* in one of three reading frames. Cells positive for YFP fluorescence were sorted using flow cytometry into 384-well plates and grown into cell clones. This screen was gated to preferentially include highly expressed proteins. Tagged protein identities were determined by 3'-RACE (rapid amplification of cDNA ends), using a nested polymerase chain reaction (PCR) that amplified the section between *EYFP* and the polyA tail of the mRNA of the host gene. PCR products were sequenced directly and aligned to the genome. Cycloheximide treatment of three clones (expressing CD-tagged USP7, HMGA2 and TOP1) indicated that the proteins retained their long lifetime (>24 h) with YFP tagging (not shown). Our library of CD-tagged proteins is detailed at www.dynamicproteomics.net.

Flow cytometry. Cells from exon-tagged clones were labelled with the vital stain SYTO-62 (Molecular Probes), which served as an indication of cell-cycle stage. Fluorescence from each cell was detected at 535 nm (YFP) and 630 nm (SYTO-62) using a Becton Dickinson FACS SORT flow cytometer (BD Biosciences). Cells at a similar cell-cycle stage (bin size of about 5% of the population by SYTO-62 fluorescence) were used for Fig. 1f.

Time-lapse microscopy. We randomly chose 20 nuclear proteins from our CD-tagged library to study variability dynamics. Time-lapse movies were obtained at 20 $\times$ magnification as described²⁴ with an automated, incubated Leica DMIRE2 inverted fluorescence microscope and an ORCA ER cooled CCD camera (Hamamatsu Photonics). The system was controlled by ImagePro5 Plus (Media Cybernetics) software which integrated time-lapse acquisition, stage movement, and software-based auto-focus. During the experiment, cells were grown and visualized in 12-well coverslip bottom plates (MatTek) coated with fibronectin (Sigma).

Image analysis of time-lapse movies. Custom image analysis software described in our previous study²⁴ performed cell tracking and segmentation, and background and bleaching corrections. Watershed segmentation was applied to images after flat-field correction and background subtraction. Tracking of cells was performed by analysing the movie from end to start and linking each segmented cell to the cell in the previous image with the closest centroid. Cell divisions were automatically detected by a sharp twofold drop in total fluorescence between consecutive images.

Auto-correlation calculation. The auto-correlation function was computed with a robust estimator $A(\tau) = \langle \langle X_i(t) \cdot X_i(t + \tau) \rangle \rangle_t / \langle \langle X_i(t)^2 \rangle \rangle_t$, where averages over time and over cells are denoted $\langle \rangle_t$ and $\langle \rangle_p$ respectively. We measured the auto-correlation of the ranked total fluorescence $R_i(t)$, such that $X_i(t) = R_i(t) - \langle R_i(t) \rangle$. This method was less sensitive to outliers in the data than auto-correlation over the total fluorescence (see Supplementary Information).

Correlations of tagged protein pairs. *mCherry*²⁹ was introduced into the genome on a retroviral vector identical to the *YFP* vector, but with *mCherry* in reading frame 0 instead of *EYFP*. Cells with YFP-tagged proteins were super-infected with this vector and sorted for red fluorescence to produce double-labelled cell clones.

Red-tagged proteins were identified with 3'-RACE (for details see Supplementary Information). About 1,000 double-labelled clones were screened for ribosomal or nuclear localization patterns, to obtain the clones studied here. Nuclear fluorescence levels of both colours were obtained from snapshots of cells, defining the nucleus by means of Hoechst 33342 (5 ng ml⁻¹) staining of DNA. Spill-over of emission from Hoechst 33342 into the YFP or mCherry channels and between tagged proteins in the same cell was negligible (<0.1% of total signal). 3'-RACE with both YFP and mCherry primers revealed that different alleles of RPL5 were labelled, with both integrations in intron-1. Normalized mCherry and YFP fluorescence distributions for RPL5 were not significantly different (Kolmogorov–Smirnov test, $P = 0.75$).

Received 8 May; accepted 2 October 2006.

Published online 19 November 2006.

- Spudis, J. L. & Koshland, D. E. Jr. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
- Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
- Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC Press, Boca Raton, Florida, 2006).
- McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
- Ferrell, J. E. Jr & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280**, 895–898 (1998).
- Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
- Bar-Even, A. et al. Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38**, 636–643 (2006).
- Newman, J. R. et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
- Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
- Austin, D. W. et al. Gene network shaping of inherent noise spectra. *Nature* **439**, 608–611 (2006).
- Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
- Mihalcescu, I., Hsing, W. & Leibler, S. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430**, 81–85 (2004).
- Colman-Lerner, A. et al. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).
- Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2**, e137 (2004).
- Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
- Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228–232 (2005).
- Volfson, D. et al. Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864 (2005).
- Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102**, 3581–3586 (2005).
- Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
- Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
- Sigal, A. et al. Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nature Methods* **3**, 525–531 (2006).
- Jarvik, J. W. et al. In vivo functional proteomics: mammalian genome annotation using CD-tagging. *Biotechniques* **33**, 852–4, 856 858–60 (2002).
- Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
- Clyne, P. J., Brotman, J. S., Sweeney, S. T. & Davis, G. Green fluorescent protein tagging *Drosophila* proteins at their native genomic loci with small P elements. *Genetics* **165**, 1433–1441 (2003).
- Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
- Shaner, N. C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22**, 1567–1572 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Kahn Family Foundation and the Israel Science Foundation for support. We thank J. Paulsson, J. Pedraza and A. Eldar for discussions of the manuscript, and A. Sharp and E. Ariel for flow cytometry assistance. R.M. thanks the Horowitz Complexity Science Foundation for support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to U.A. (uri.alon@weizmann.ac.il).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Maybe the prospect of a big meal and a few days off breached my professional cynicism, but last week's US Thanksgiving holiday gave me pause to reflect that, in the six or so years that I've been writing this column, scientists have accrued a number of reasons to be thankful. I'll start in the United States. Although the budget for the National Institutes of Health has been somewhat flat in recent years, it is still much larger than it was six years ago. And more money is being spent on training schemes, such as the \$75 million recently announced for training in bioinformatics research. I must admit, however, that many scientists might also view politics as obstructive to some research areas. That would probably be the general feeling about the US policy that restricts funding for research on human embryonic stem cells. But even on this contentious issue, individual states and foundations have taken the initiative — take the recent referendum in Missouri, which provided a narrow victory in favour of state funding for stem-cell research.

Much of the world might not celebrate Thanksgiving, but scientists outside the United States have some things to be grateful for, too. The growth of major biomedical research foundations, such as the Wellcome Trust, the Bill & Melinda Gates Foundation and the Howard Hughes Medical Institute, means that billions of dollars are being funnelled into unique approaches for doing research globally. Students and postdocs, meanwhile, are forming more organizations and associations to help young scientists explore their career options. And many European countries have agreed to increase their financial commitment to research and development, and are working to build bridges between academia and industry.

As for me, on my travels throughout North America, Europe and Asia, I'm grateful for numerous acts of kindness — from countless cups of coffee and tea, to a recent ride on the back of a bicycle that helped me get to a key interview. But the biggest debt of thanks I owe is to the hundreds of scientists who have agreed to speak to me about their career issues and aspirations in order to help their colleagues. Thanks, everybody.

Paul Smaglik, *Naturejobs* editor

Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestranguez (4994)

UK/Ireland/Italy/RoW:
Loredana Milanese (4944)
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

European Satellite Office
Germany: Patrick Phelan
Tel: +49 89 54 90 57 11
Fax: +49 89 54 90 57 20
e-mail: p.phelan@nature.com

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production:
Catherine Alexander

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com



BIO BONANZA

Is interest in biofuels in the United States a just fad or a growing trend that will yield numerous jobs and research opportunities? **Gene Russo** separates the wheat from the chaff.

Scott Bents likes to design things. So after receiving his bachelor's degree in mechanical engineering from Iowa State University, he began designing machine parts at a washing-machine factory. Five years later, his job moved to Canada and the factory shut down.

Unemployed, Bents turned to a government programme that, as part of the 1993 North American Free Trade Agreement, funds retraining for workers whose jobs have gone to Canada or Mexico. Seeking a rewarding career and a worthy cause, he settled on biofuels research — and the promise of obtaining renewable and sustainable energy from biomass such as the native plant switchgrass. Now, as part of his PhD in mechanical engineering at Iowa State, Bents is trying to design an economically viable biomass refinery that produces hydrogen and a biodegradable plastic. He is one of many scientists keen on the potential of biofuels.

His story embodies the hope of farmers, agribusinesses, US heartland politicians and biofuels researchers and enthusiasts. They envisage a day when the biofuels industry will revitalize towns of the depressed Midwest, brought to life with a new jobs sector and the hope of a renewable energy source to slake the country's insatiable thirst for fuel.

Iowa and other states are already on the brink of a corn ethanol boom. Ninety-seven corn ethanol plants currently operate in the United States and dozens more are under construction, according to the Renewable Fuels Association trade group. It's big business, made quite profitable by a government ethanol subsidy (51 cents per gallon) and a ban on methyl *tert*-butyl ether (MTBE) as an additive to unleaded petrol. A toxic pollutant, MTBE is being replaced by ethanol as the additive of choice to help petrol burn more cleanly.

These factors have aroused plenty of industry interest in corn ethanol refineries and, potentially, could lead to more jobs; right now, most jobs being created are for plant operators and don't require science training. But many see cellulosic ethanol as the logical

renewable choice of the future, a potential boon for researchers and engineers.

Chemist Art Ragauskas, a cellulosic ethanol researcher at Georgia Institute of Technology, Atlanta, was excited when he heard President Bush's State of the Union address last January. Bush not only asserted that the United States was "addicted to oil", says Ragauskas, but he actually uttered the word "switchgrass". Ragauskas began his career in the pulp and paper industry before applying his knowledge of biomass polymers to biofuels. He has seen a major surge in interest since he entered the field six years ago.

Powerful force

This is not the first time cellulosic ethanol has received attention. Interest piqued in the 1970s when oil prices skyrocketed. When the prices dipped again, so did enthusiasm for ethanol. But researchers and policy-makers see a perfect storm of political attention, rising oil prices, fear of environmental impacts from fossil fuels, and new genomics tools that can modify plants and microbes. And, they say, cellulosic ethanol research is poised for a major windfall.

"Cellulose and biomass have the potential to create a great deal more energy than corn, and with each day we come closer to a cost-effective technique for releasing that energy on a broader scale," Mike Johanns, secretary of the US Department of Agriculture (USDA), told a renewable-energy conference in St Louis in October.

Corn ethanol, derived from the simple sugars of the kernel, requires lots of land, water and groundwater-tainting fertilizer to produce. Cellulosic ethanol biorefineries, still only at the earliest pilot stage, are widely considered preferable. They could achieve much greater energy return and environmental sustainability by using corn stover (the stalks, husks and leaves that are normally discarded), switchgrass, poplar trees and other biomass sources. Scientists and engineers are seeking a fast and economical way to penetrate the plant cell's stubborn cell wall, use enzymes to break



Robert Brown: seeing an increase in student demand.

down cellulose into its component sugars, and convert those sugars into ethanol with the help of microbes.

Already the federal government has shown some commitment to biofuels. As dictated by the US Energy Policy Act of 2005, the United States aims to replace 30% of its transportation fuel with biofuels by 2030. Only half of that goal can be met by corn ethanol, says Brian Davison, director of the bioprocessing research and development centre at Oak Ridge National Laboratory. Cellulosic ethanol could help fill the gap. Although the government's 2007 budget has not been finalized, the House's budget requests \$150 million for biomass research and the Senate's \$213 million, up from \$94 million from 2006.

In August, the US Department of Energy (DOE) requested proposals for two new bioenergy research centres for a total of \$250 million, as part of its 'genomes to life' programme. Universities, national laboratories, nonprofit organizations and private firms can compete for the awards, which will be announced in summer 2007. "Everyone in the field is trying to find a way to position themselves for one of these bioenergy centres," says Kevin Chambliss, an analytical chemist and cellulosic ethanol researcher at Baylor University in Texas. The centres will conduct systems-biology research on microbes and plants to optimize the conversion of cellulose into fuel. The DOE and USDA recently awarded \$17.5 million for 17 biomass research, development and demonstration projects, and Chambliss says several states are awarding modest amounts for biofuels research; last year his group received \$250,000 from the Texas government.

Considering the size of the challenge, some feel that the government's contribution has fallen short. "The percentage increase may seem dramatic, but it's not a whole lot of money," says Reid Detchon, executive director of the nonprofit Energy Future Coalition. Quang Nguyen, a programme manager at Spanish company Abengoa Bioenergy's site in Chesterfield, Missouri, suggests that the DOE has done a good job of providing focused funding to push key technologies, such as pretreatment processes and cellulose enzymes.

Government funds have been augmented by some investment from industry. In June, Chevron formed an alliance with the Georgia Institute of Technology. Chevron will contribute up to \$12 million over five years in an effort to make cellulosic ethanol commercially viable. BP has pledged \$500 million over



Scott Bents: aiming to design a biorefinery.

ten years to establish a biosciences energy research laboratory, focused largely on cellulosic ethanol, to be attached to an academic centre in the United States or Britain. Candidates are still being considered; the research programme will be launched by late 2007. A \$38-million consortium of the DOE's National Renewable Energy Laboratory, DuPont, Diversa and Michigan State University, started in 2003, is attempting to develop the optimal biorefinery.

"What makes me most confident in the potential of renewable energies is the market investment we've seen," said Johanns at the October meeting. For cellulosic ethanol science and technology to reach its potential, the field will require a mix of disciplines: chemical and biochemical engineers to develop the processes and equipment, and microbiologists to design efficient microbes and enzymes. There will also be a demand for technicians, biorefinery plant operators and lab scientists to maintain quality control.

Just add (very little) water

Agricultural researchers and geneticists, meanwhile, are looking to boost crop yields and design plants tailor-made to maximize ethanol potential. Nguyen envisages merging marginal lands with specially bred or engineered plants to create 'energy plantations'. The St Louis-based seed company Monsanto is developing an experimental drought-resistant corn variety that, so far, has given increased yields of 8–10%. Chief technology officer Robert Fraley suggests that yield could double in the next 30 years. Pradip Das, Monsanto's director of crop analytics, says the company is considering hiring scientists who have backgrounds suited to ethanol research. Ideally, he says, candidates would have a biology background as well as some knowledge of the ethanol industry and the engineering processes involved.

Universities are taking an interest as well (see 'A biofuelled expansion'). At Iowa State, Bents is completing a new specialization in biorenewable resources and technology. "Colleagues at other schools had cautioned us against doing this programme," says Robert Brown, director of the programme. "Now we're seeing an increase in demand." Programmes should be interdisciplinary, says Brown, including fundamental courses such as organic chemistry and economics, as well as topics such as environmental impact, agronomics, plant science, processing and production.

Students, says Ragauskas, often get excited about work that could benefit the environment. "Students always want to do something that makes a difference," he says. Ragauskas once actively recruited PhD students and postdocs. Now he has a hard time finding postdocs, something he attributes, in part, to wide opportunities in industry. Chemical engineer and cellulosic biomass researcher Y. Y. Lee, at Auburn University in Alabama, says his students have no problems finding positions. Two of his undergraduates have secured jobs at Exxon and Novozymes North America in Franklinton, North Carolina. His recent graduate students have gone on to take positions at Abengoa Bioenergy, Novozymes and a postdoc position at the USDA.

Bents, now in his third year, is still deciding between academia and industry. If oil prices stay high, federal funding continues to increase and industry continues to invest, he should have plenty of options.

Gene Russo is assistant editor of *Naturejobs*.

A BIOFUELLED EXPANSION

With researchers, policy-makers and prospective students showing an increasing amount of interest in the potential of biofuels, US universities are attempting to build or expand academic programmes.

Long-time cellulosic ethanol researcher Lee Lynd, at Dartmouth College in Hanover, New Hampshire, has seen a major surge in recruiting interest from universities in the past 12 months. "It's a complete about-face," says Lynd.

Dartmouth, Iowa State, Michigan State, North Carolina State and Purdue universities, as well as the University of Georgia, are among those institutions that offer coursework, certificate programmes

and/or specializations that are related to biofuels research.

During the past four years, Purdue's agricultural and biological engineering programme has increased its number of biofuels-related faculty members, says department head Bernie Engel. And North Carolina State's biological and agricultural engineering department offers a bioprocessing track focused on the study of biofuels and other bioproducts.

"I have more students interested in my research programme than I have money to fund them," says Mari Chinn, a North Carolina State assistant professor of bioprocessing and engineering. G.R.

The charge-up man

Last delivery before the Singularity.

Catherine H. Shaffer

My charge-up man came on the very last day before the Singularity. There were butterflies in my stomach as I heard the music from a distance. Then his van crested the hill. His van, its sides painted all in red and blue, like an old-time gypsy caravan. Children came running from down the street, with phones and music pods and game packs in their little hands. They laughed when the charge-up man touched their devices to his charger. They should work for a whole month, now. A month!

Of course, the whole world was going to change in 24 hours. We'd been getting ready for the Singularity for so long, I couldn't quite believe that it was finally going to happen.

The government men had come around with their pamphlets. They handed out iodine tablets and duct tape and air freshener and other things that didn't make any sense. The rate of technological progress was going to go exponential, they said. Computers and artificial intelligences and nanotech devices were going to, at some point, get up and start improving themselves, and after that, the world would change in ways that no human being could imagine. Then they left.

The charge-up man was different. He'd started coming around when our stuff stopped working. Software updates would fail, or power adapters would become obsolete. Or sometimes the whole block would brown out, and he would park his van with its generator and go around powering everybody up again. The Singularity needs a lot of energy.

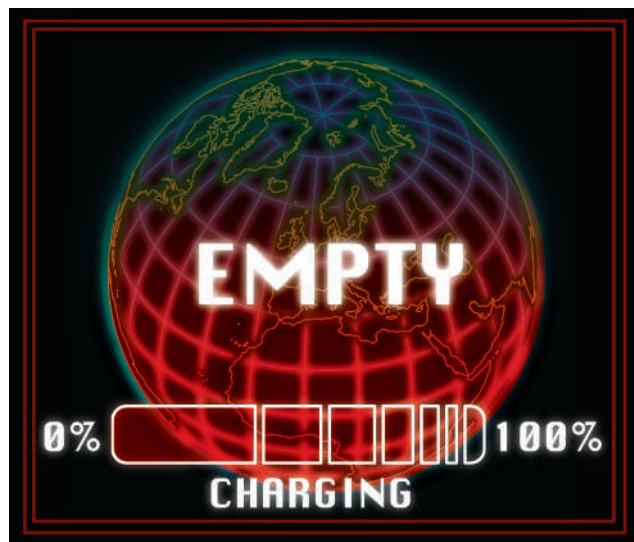
I had to fan myself a little with my newspaper as I went to the front door to open it. My charge-up man was really cute. I opened the door and he smiled at me. "Charge up, Ma'am?"

I stepped aside and let him in. I was wearing my low-cut top and I had put on make-up and perfume. Just for him. I was feeling reckless. Everybody was, that day.

My charge-up man wasn't the right man for me. We had never even traded profiles. But what the heck? A day later, we might find ourselves in a travelling mariachi band, or living underwater with tentacles growing out of our heads. Nobody knew what was

going to happen after the Singularity.

The charge-up man moved around my apartment, holding his charger, with all of his adapters and cords on a belt around his hips. He had wavy hair with little curls on the back of his neck that I wanted to twirl with my fingers.



He charged up all my newspapers, magazines and books. He charged up my phone and my pager and my toothbrush. He upgraded the cat and the catfood dish and the vacuum cleaner. The charge-up man had a list of all the devices in my house that needed charging and upgrading. I couldn't keep track of them. I remembered the old days, when we had to charge our own stuff, and the battery only lasted a few days, or a few hours, instead of a whole month. It was so inconvenient. And I got so angry when I wanted to use my phone or my pocket computer and tried to turn it on, only to get a warning: *Low battery. Please shut down immediately to avoid information loss.*

"What's your name?" I asked him.

He was squatting, charging up the little dinosaur that washes my floors. I bent over so he could see down my shirt when he looked up.

"My name is Brian. What's yours?" He was definitely looking down my shirt.

"It's Rita."

"Hi, Rita," he met my eyes for a moment, and then looked at the dinosaur again. "I don't have the right adapter for this device. I'll need to get a larger male adapter from my van. I'll be back in a minute."

Too soon, Brian was done charging up everything in the house, and it was time for him to go.

As he stood at the door, hesitating slightly, I gathered up my courage. "Couldn't you stay?"

He smiled at me. "Not now," he said. "Everyone wants a charge-up with the Singularity coming tomorrow."

"Then I won't see you again until...?"

Brian was framed in the doorway, sunshine bouncing off reddish-blond hair. He flashed a smile at me. "I'm not working tomorrow."

I beamed back at him. "Perfect."

After he left, I collapsed in a heap on my sofa, my heart flipping over in my chest.

Singularity day was a worldwide holiday. All my books and newspapers had stopped working in the night. My toaster had a core dump and wouldn't boot up, so I had a bowl of cereal instead.

Brian came over and we went outside. Wispy clouds were drifting across a bright blue sky. My house had com-

pletely shut down, and I had a message that everything in it was obsolete, and would be upgraded soon. A crowd made up of my confused neighbours milled around in the street.

Someone had a watch, and started counting down.

We joined in: "Seven, six, five, four, three, two..."

Brian squeezed my hand as we both said, "One!"

There was silence. After a long minute, a child asked: "What happened?"

I sat talking with Brian later. People had brought out lawn chairs and lit barbecue grills. They were still waiting for something to happen. "I thought the Singularity would be more impressive," I said, as I slid a marshmallow onto the end of a stick.

Brian shrugged. "A Singularity needs a lot of energy." He took out his PDA and thumbed it on. The screen lit up briefly, then a message came up before it dimmed again: *Low battery. Please shut down immediately to avoid information loss.*

Catherine Shaffer is a science and science-fiction writer from southeast Michigan. She is a frequent contributor to *Analog* and has published stories in many other magazines and anthologies. She can be found on the web at www.CatherineShaffer.com.

JACEY